





ACCOUNTABILITY AND ASSESSMENT

ASSESSMENT DATA CAN PROVIDE A WIDE RANGE OF INFORMATION THAT WILL HELP STATES, DISTRICTS, AND SCHOOLS DEMONSTRATE THEIR ACCOUNTABILITY IN TERMS OF THEIR EDUCATIONAL RESPONSIBILITIES TO CHILDREN, PARENTS, AND COMMUNITIES. A CORE ELEMENT OF READING FIRST AND EARLY READING FIRST IS REGULAR, SYSTEMATIC ASSESSMENT OF CHILDREN'S READING ABILITY TO DETERMINE THEIR READINESS FOR INSTRUCTION, TRACK THEIR PROGRESS OVER TIME, AND PREDICT THEIR LIKELY SUCCESS IN THE FUTURE.

THIS SECTION OF THE GUIDEBOOK INCLUDES:

-  A PowerPoint presentation on accountability issues related to reading
-  A PowerPoint presentation on the dimensions of assessment important to reading instruction
-  An Action Plan for Implementing a District-wide Early Assessment System
-  The Importance and Decision-making Utility of a Continuum of Fluency-based Indicators of Foundational Reading Skills for Third Grade High Stakes Outcomes: An Introductory Paper

A POWERPOINT PRESENTATION ON ACCOUNTABILITY
ISSUES RELATED TO READING



Reading First Academy Assessment Committee

Team Leader Edward J. Kame'enui, University of Oregon

- **David Francis, University of Houston**
- **Lynn Fuchs, Vanderbilt University**
- **Roland Good, University of Oregon**
- **Rollanda O'Connor, University of Pittsburgh**
- **Deborah Simmons, University of Oregon**
- **Gerald Tindal, University of Oregon**
- **Joseph Torgesen, Florida State University**

Assessment in the Context of State Accountability

- Assessment can provide a vivid, public report of reading outcomes.
- Assessment can inform instruction to help schools, teachers, and children achieve important reading outcomes.
- Accountability is crucial to the educational agenda.
 - 31 states use norm-referenced tests for accountability, 33 use criterion referenced tests, and 19 use both (Bond, Rueber, & Connealy, 1998).
 - Accountability for reading outcomes is central to *Reading First Initiative*.

- Only 5 states do not use criterion-referenced or norm-referenced tests for accountability.
- Good, R. H., Simmons, D. C., & Kame'enui, E. J. (2001). The importance and decision-making utility of a continuum of fluency-based indicators of foundational reading skills for third-grade high-stakes outcomes. *Scientific Studies of Reading*, 5, 257-288.

Accountability for Grade Level Reading Outcomes

Frequently, statewide accountability assessments are given only at selected grades, so accountability for grade level reading outcomes raises two questions:

- *Do students meet state standards on the current state assessment?* Students are *at grade level* if they meet expectations for reading proficiency on a state accountability assessment.
- *Are students predicted to meet state standards on the next state assessment?* Students are *at grade level* if they are likely to meet the state standard on a later accountability assessment (assuming effective comprehensive reading instruction).

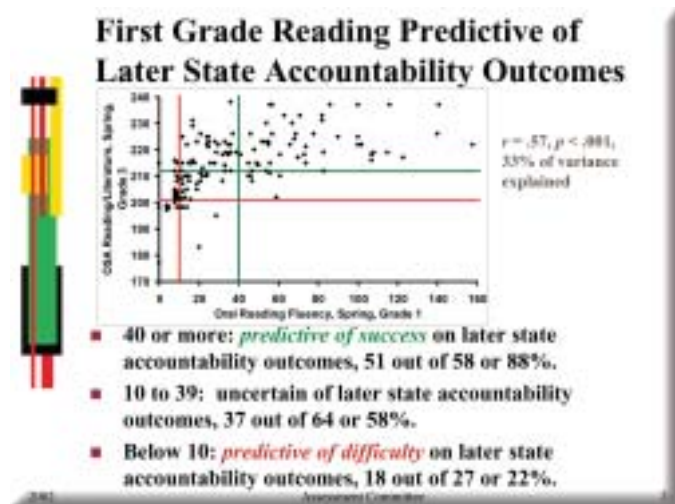
- If a state accountability assessment isn't given until third grade, for example, then we need some way of finding out if children are on track for that accountability assessment.
 - First, it makes no sense to place all of the stress and responsibility on the third grade teachers only when third grade outcomes are the culmination of K through third grade instruction.
 - Second, waiting until third grade to assess outcomes and accountability limits our ability to do something about it. At kindergarten, first, and second grade we need accountability for progress and outcomes in time to make adjustments in instruction to change third grade outcomes.
- So, even if the first formal state accountability assessment does not occur until third or fourth grade, we need intermediate accountability assessments to show that children are on track and are likely to meet the state accountability assessment.

Example: Oregon Statewide Assessment in Reading and Literature

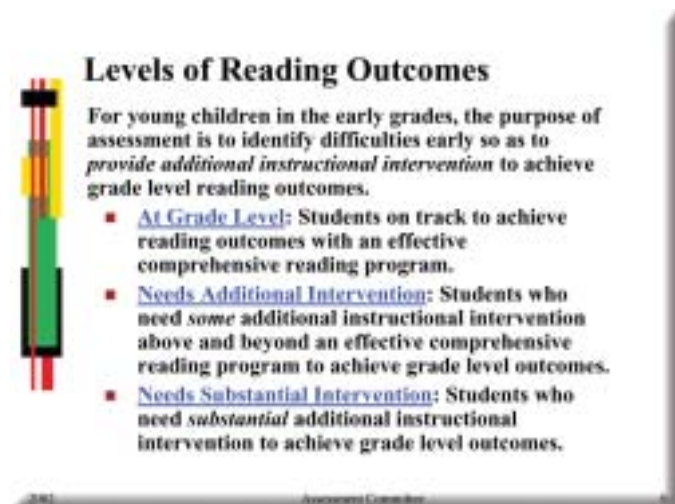
- The Oregon Statewide Assessment in Reading and Literature (OSA) is a standardized reading achievement test developed by a panel of experts and a research and development company.
- Standards are established by the Oregon State Board of Education.
 - In third grade, a score of *201 or higher* is judged to "*meet expectations*" and a score *below 201* is described as "*does not meet expectations*."
- The OSA is repeated in grades 3, 5, 8, and 10.
 - In third grade, a score of *212 or higher* is *predictive of success* on the 5th and 8th grade state accountability assessments.

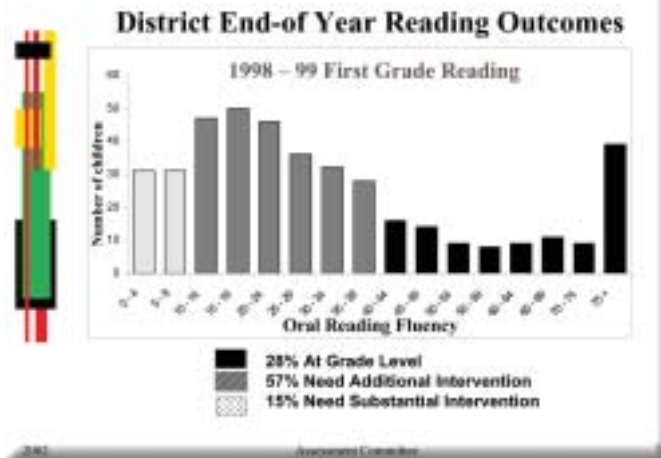
- Longitudinal research can examine the predictive validity of earlier grade accountability goals for later grade state accountability assessments.
- Oregon is used here only as an example. Similar research and conclusions can be expected for other state accountability assessments.
- Two questions are important: Are students meeting state accountability standards now? And, are they on track to meet later state accountability standards.
- For example, what about a student who receives a score of 202 on the OSA in third grade? That student would be rated as "meets expectations" for third grade, but they may not be rated as "meets expectations" on the fifth grade and eighth grade assessments.

- In first grade, a strong prediction that a student is on track for the third grade state accountability assessment is possible.
- No prediction is ever perfect -- a perfect prediction would mean that second and third grade instruction did not matter in changing outcomes. In first grade, the goal is to reach a level of reading skill where the odds are that the child will meet state accountability standards.
- Reading the picture (optional discussion). Each dot represents a child. If you go straight down from the dot, the score represents the child's first grade performance. If you go straight left from the dot, the score represents the child's third grade performance on the state accountability assessment.
 - If you cover all the dots for children who scored below 40 in first grade (to the left of the green line) and look only at the children who met the first grade goal, all met the third grade state accountability standard. And, 88 percent were predicted to meet the fifth and eighth grade state accountability standard.
 - If you cover all the dots for children who score above 10 in first grade (to the right of the red line) and look only at the children who were predicted to experience difficulty on the state accountability standard, only 22 percent were predicted to meet fifth and eighth grade state accountability standard.

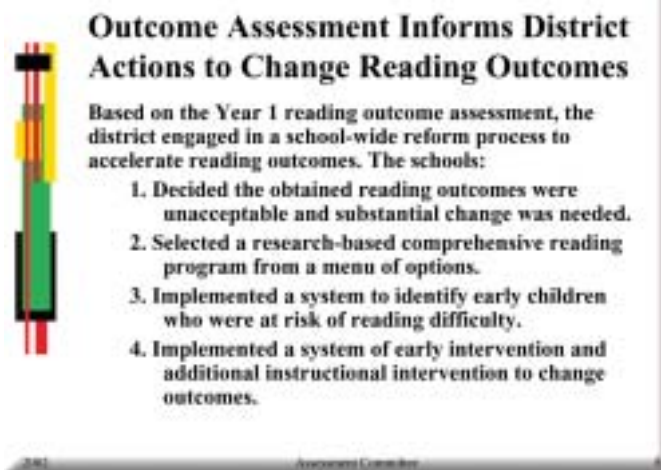


- Especially in the early grades, one purpose of accountability assessment is to provide intervention to change the prediction. Predictions of difficulty on later state accountability assessments are only accurate if we don't do anything to change the outcome.
- A second purpose of accountability assessment at each grade is to inform decisions about curriculum, instruction, and intervention to get more children on track for later state accountability assessments.

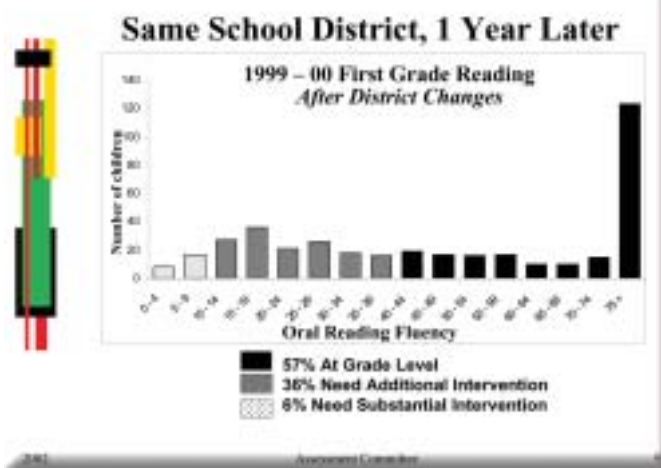




- This slide and the next three slides tell the story of a district that used first grade outcome assessment to make decisions about their curriculum, instruction, intervention, and allocation of resources.
- In the first year, 1998-1999, they found that only 28% of their children were on track for success on the 3rd, 5th, and 8th grade state assessments. For 57% a clear prediction was not possible, and for 15% difficulty passing was predicted unless substantial intervention was provided.



- The first grade accountability assessment was only one piece of the change process, but it provided a clear basis for a decision that a change was needed, and it provided a clear goal for the reform process.



- In the second year, 1999-2000, for their next cohort of first graders, the results of the district's work and effort to reform their curriculum, instruction, intervention, and allocation of resources was clearly and vividly apparent. They doubled the percent of children at grade level, and they halved the number of children predicted to experience difficulty on the state accountability assessment without substantial intervention.
- Comparing Slide 9 with Slide 7 illustrates the clear difference in outcomes achieved by the district.
- For this district, accountability for outcomes provided a basis for celebration -- and renewed effort.

- Of course, it is possible that the change was due to a different group of students or to a change in the district demographics. By including a measure of beginning skills for each group of students and by examining the overall pattern of change over time, other explanations for changes in outcomes can be examined for the district.

Outcome Assessment Informs District Decisions and Actions

Based on the Year 2 reading outcome assessment, the district decided:

- They were making dramatic changes in the reading outcomes for children in their district.
- Renewed effort was still needed, including continued professional development to implement a comprehensive reading program with increasing fidelity.
- They needed to institute a system of increasing the intensity of instructional interventions for students who require more intensive intervention to be at grade level.

- Accountability assessment should focus on outcomes in the important beginning reading areas. The important beginning reading areas are the ones where we can make the most difference in reading outcomes for children.
- All beginning reading areas are not equally important at different ages. For example, foundation skills are more important early -- reading comprehension skills are more important later.


Reporting Outcome Assessment

- Outcomes in the important beginning reading core areas *should be reported* at the beginning and end of the academic year.
 - Phonemic Awareness
 - Phonics
 - Fluency
 - Comprehension
 - Vocabulary
- The following assessment framework specifies the important beginning reading skill areas to be assessed at the beginning and end of each grade level.

- These charts show a recommended minimum assessment schedule. Districts may assess more frequently or in more skill areas, but an assessment of phonemic awareness and vocabulary at the beginning and end of kindergarten and a measure of phonics at the end of kindergarten are suggested.

Kindergarten Outcome Assessment

	Suggested Beginning of year	Optional Middle of Year	Suggested End of Year
Phonemic Awareness	X	X	X
Phonics		X	X
Fluency			
Comprehension			
Vocabulary	X	X	X




First Grade Outcome Assessment

	<i>Suggested Beginning of year</i>	<i>Optional Middle of Year</i>	<i>Suggested End of Year</i>
Phonemic Awareness	X	X	X
Phonics	X	X	X
Fluency		X	X
Comprehension			X
Vocabulary	X	X	X

2001 Assessment Committee 13

- In later grades, once children have foundation reading skills, the focus of assessment shifts to fluency and reading comprehension.




Second Grade Outcome Assessment

	<i>Suggested Beginning of year</i>	<i>Optional Middle of Year</i>	<i>Suggested End of Year</i>
Phonemic Awareness			
Phonics	X		
Fluency	X	X	X
Comprehension	X	X	X
Vocabulary	X	X	X

2001 Assessment Committee 14

- For most children, assessment of phonemic awareness and phonics skills is less important in later grades.




Third Grade Outcome Assessment

	<i>Suggested Beginning of year</i>	<i>Optional Middle of Year</i>	<i>Suggested End of Year</i>
Phonemic Awareness			
Phonics			
Fluency	X	X	X
Comprehension	X	X	X
Vocabulary	X	X	X

2001 Assessment Committee 15

- To be maximally useful, accountability assessment should be focused and clear. The purpose is not to overwhelm users and consumers with a morass of scores, subtests and numbers. For the example district illustrated earlier, the first year they had 26% at grade level at the end of first grade. The next year they had 57% at grade level.
- Providing a clear goal for each grade and providing clear information about progress toward the goal can be an important support for a district in changing reading outcomes and meeting state accountability assessments.




Suggested District-Level Reading Outcome Summary

- Overall, end-of-year, number and percent of children who are:
 - At grade level
 - Need additional intervention
 - Need substantial intervention
- Percent of children who are at grade level by racial and ethnic status, and by free/reduced lunch status as an indicator of low income.
- Change in percent of children who are at grade level compared to:
 - Beginning of year percent at grade level
 - Prior year percent at grade level

2012 Assessment Committee

- Accountability or outcome assessment should be one piece of an assessment system. All pieces of the system need to be in place and need to work together to support reading outcomes for our children. Other pieces of a complete assessment system will be discussed in the assessment presentation.



Four Kinds of Reading Assessments

An effective, comprehensive, reading program includes reading assessments to accomplish four purposes:

- *Outcome* - Assessments that provide a bottom-line evaluation of the effectiveness of the reading program.
- *Screening* - Assessments that are administered to determine which children are at risk for reading difficulty and who will need additional intervention.
- *Diagnosis* - Assessments that help teachers plan instruction by providing in-depth information about students' skills and instructional needs.
- *Progress Monitoring* - Assessments that determine if students are making adequate progress or need more intervention to achieve grade level reading outcomes.

2012 Assessment Committee

A POWERPOINT PRESENTATION ON THE
DIMENSIONS OF ASSESSMENT IMPORTANT TO
READING INSTRUCTION



Reading First Academy Assessment Committee

Team Leader Edward J. Kame'enui, University of Oregon

- **David Francis, University of Houston**
- **Lynn Fuchs, Vanderbilt University**
- **Roland Good, University of Oregon**
- **Rollanda O'Connor, University of Pittsburgh**
- **Deborah Simmons, University of Oregon**
- **Gerald Tindal, University of Oregon**
- **Joseph Torgesen, Florida State University**

Four Kinds of Reading Assessments

An effective, comprehensive, reading program includes reading assessments to accomplish four purposes:

- **Outcome** - Assessments that provide a bottom-line evaluation of the effectiveness of the reading program.
- **Screening** - Assessments that are administered to determine which children are at risk for reading difficulty and who will need additional intervention.
- **Diagnosis** - Assessments that help teachers plan instruction by providing in-depth information about students' skills and instructional needs.
- **Progress Monitoring** - Assessments that determine if students are making adequate progress or need more intervention to achieve grade level reading outcomes.

- In the Accountability section, we discussed the importance of accountability assessment to provide a clear goal and our year to year progress toward the goal. Accountability or outcome assessment provides the cornerstone for an assessment system that can support schools to change reading outcomes for our children.
- Screening assessment is used to identify children early in the school year who may experience difficulty meeting standards on the end of year accountability or outcome assessment.
- Diagnostic assessment informs a specific instructional plan for what we need to do to change the outcome and to ruin the prediction of difficulty.
- Progress monitoring assessment provides continuous, ongoing, formative information that is used to evaluate and modify the instructional plan.

Screening Assessment

- The crucial issue for screening assessment is predictive validity - which children are likely to experience reading difficulty?
- The primary purpose of screening assessment is to identify children early who need additional instructional intervention.
- *Identification is not enough!* Screening is only valuable when followed with additional instructional intervention so that students achieve grade level reading outcomes.

- Screening should be closely aligned with the outcomes and accountability assessment. A crucial role of screening is to identify children who may not reach the outcome or accountability standard unless we provide additional intervention. Important beginning reading core areas include:
 - Phonemic Awareness
 - Phonics
 - Fluency
 - Comprehension
 - Vocabulary
- Screening must fit within an assessment system that targets important outcomes, and leads to the development of an intervention plan to change those outcomes for students at risk.

National Reading Panel. (2000). Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction: Reports of the subgroups. Bethesda, MD: National Institute of Child Health and Human Development.

- This slide and the next 4 slides tell the story of five elementary schools that combined screening for children at risk with substantial intervention support to change reading outcomes and subvert the prediction of reading difficulty for the children screened as at risk. In this longitudinal study, 201 randomly selected children from five elementary schools serving children from mixed SES and ethnic backgrounds (28% free and reduced lunch) were followed from the beginning of first grade through the end of fourth grade.
- For kindergarten, screening is focused on phonemic awareness, the key accountability assessment outcome.
- First, the researchers followed the children identified as at risk to see if the prediction of risk was accurate.

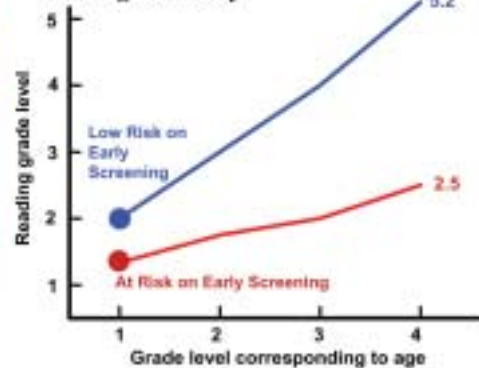
Early Screening Identifies Children who Need Additional Intervention

For example, in one longitudinal study:

- 201 randomly selected children from five elementary schools serving children from mixed SES and ethnic backgrounds were followed from the beginning of first grade to the end of fourth grade.
- Children who scored low on phonemic awareness and letter knowledge at the beginning of first grade
 - Started with lower skills
 - Made less progress
 - Fell further and further below grade level as they progressed from first through fourth grade.

- This slide demonstrates the power of screening assessment to predict reading outcomes through the end of fourth grade. Reading outcomes were measured at the end of each year. The measure reported here assesses a combination of reading accuracy and comprehension.
- The children were administered measures of phonemic awareness and letter knowledge at the beginning of first grade, and divided into two groups: *At-Risk*, and *Low Risk*.
- The line in red shows the progress of children who began first grade performing in the bottom 15% in phonemic awareness and letter knowledge. At the end of fourth grade, these children were reading at an average level of mid second grade. In contrast, children who began first grade with higher levels of phonemic awareness and letter knowledge and roughly equivalent levels of overall ability, finished fourth grade reading at beginning fifth grade level.

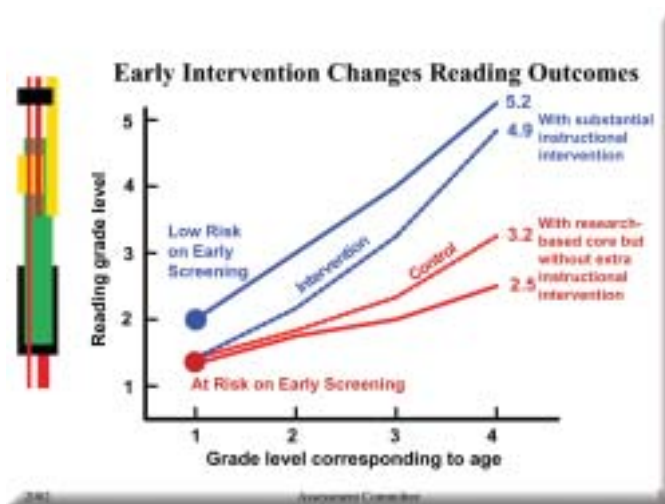
Early Screening Identifies Children At Risk of Reading Difficulty



- The four years illustrated in the previous slide were a passive observation study. The researchers identified the children at risk, but then did not provide any systematic change in instruction beyond the current practices of the school.
- The investigators who conducted this second study were able to go back to the same schools about four years later, after many of the schools had switched to a more research-based, comprehensive reading program in grades K through 3. The research-based comprehensive reading program was implemented for the entire school, including the children identified as at risk on the same screening assessment.

Additional Instructional Intervention Changes Reading Outcomes

- Four years later, the researchers went back to the same school. Two major changes were implemented:
- First, a *research-based comprehensive reading program* was implemented for all students, and
- Second, children at risk for reading difficulty were randomly assigned to a control group or to a group receiving *substantial instructional intervention*.



- Children from the bottom 15% in phonemic awareness and letter knowledge were randomly assigned to either a control group, or a group that received more intensive reading instruction in first and second grade.
- The dotted red line shows the progress of the children who did not receive extra instructional intervention, and you can see that improved classroom instruction produced slightly better outcomes for them than in the earlier study in the same schools.
- However, the children who were identified by the screening tests *and* received *substantial instructional intervention* did almost as well as average children by the end of fourth grade.
- Improved classroom instruction will help our most at-risk children learn to read better, but most will require more intensive interventions if we expect them to read at grade level by the end of fourth grade.

Research-Based, Comprehensive Reading Program and Substantial Instructional Intervention

Both a *research-based comprehensive reading program* and *substantial instructional intervention* were needed for children at risk of reading difficulty. Children receiving substantial additional instructional intervention beyond an effective comprehensive reading program:

- Progressed more rapidly than control students,
- Had reading skills more like the low risk group than the at risk group, and
- Were reading about at grade level.

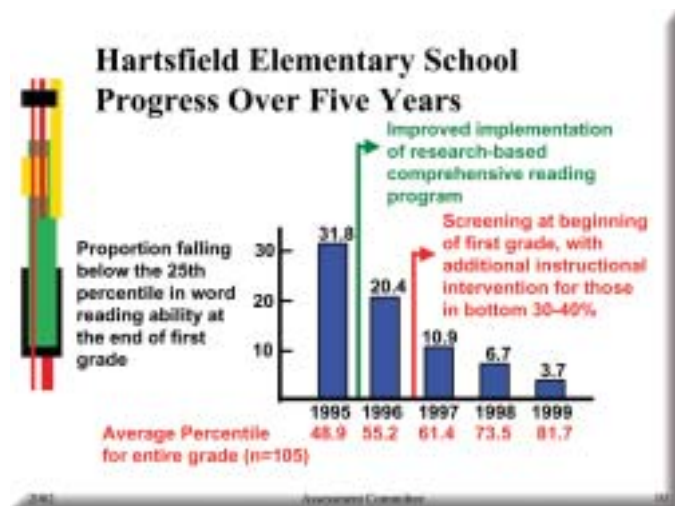
- If we are going to leave no child behind, we need **both** a research-based comprehensive reading program **and** substantial instructional intervention for students who are screened as at-risk for reading difficulty.

Improving the Reading Program by Adding Assessment and Intervention

- Hartsfield Elementary School Characteristics:**
 - 70% Free and Reduced Lunch (increasing)
 - 65% minority (mostly African-American)
- Elements of Curriculum Change:**
 - Movement to a more research-based reading curriculum beginning in 1994-1995 school year for K-2 (incomplete implementation)
 - Improved implementation in 1995-1996
- Implementation in Fall of 1996 of screening and more intensive small group instruction for at-risk students

- This slide and the next slide tell the story of Hartsfield Elementary School.
- Hartsfield modified its reading curriculum in grades K-2 with a research-based, comprehensive reading program.
- They also added intensive instructional intervention for children who were at-risk on a screening assessment. At the time these data were collected, Hartsfield Elementary School was serving a student population in which 70% of the children received free or reduced lunch, and 65% were minority students.

- In 1994, the school began to implement a more research-based, balanced reading curriculum in its K-3 classrooms. This change took them about two years to implement fully.
- After the first year of implementation of the more balanced curriculum, about 32% of the children finished first grade with reading scores below the 25th percentile.
- In 1996 the research-based, comprehensive reading program was more fully implemented.
- With the more complete implementation, about 20% were reading below the 25th percentile.
- In 1997 they began screening children at the beginning of first grade and providing intensive instructional intervention to children who were at-risk on their measures.
- The next year, only 11% of the children were still poor readers at the end of first grade.
- By the end of 1999, only about 4% of the children were reading below the 25th percentile.
- During this same period of time, the overall percentile in reading ability of all the children at the end of first grade increased from the 49th percentile at the end of 1995 to the 82nd percentile at the end of 1999.



- The purpose of Diagnostic Assessment is to provide more in-depth information on a student's skills and instructional needs to plan instruction.
- Diagnostic assessment may include a variety of teacher-made or specialized assessments, but should be closely aligned to the accountability outcomes for the grade.
- For children at risk of difficulty achieving accountability standards, which beginning reading core area are they experiencing difficulty in?
 - Phonemic Awareness
 - Phonics
 - Fluency
 - Comprehension
 - Vocabulary
- Once the beginning reading core area is identified, knowledge of specific skill difficulties are needed to inform intervention. If students are having difficulty in phonemic awareness, have they mastered initial sounds? Final sounds? Medial sounds? What point in an instructional intervention for phonemic awareness would be a good starting point for instruction?
- Diagnostic assessment aligned with outcomes and screening supports the development of good intervention plans for students who are at risk for reading difficulty.

Diagnostic Assessment For Students Who Need Additional Intervention

- On which of the important beginning reading skill *areas* are the students on track, and on which do they need additional instructional intervention?
- Which *specific* beginning reading skills has the student mastered or not mastered?
- How much instructional intervention are the students likely to need (e.g., smaller group, extra time, more practice, more modeling, more scaffolding)?
- Which intervention programs are most likely to be effective?
- Which students have similar instructional needs and will form an appropriate group for instruction?



Progress Monitoring Assessment

- Children respond differently, even to instruction that is research based and usually effective.
- If we are to get *all* children at grade level, we must get *each* child at grade level -- and keep them there.
- We need to identify early when children begin to get off track and make necessary modifications to instruction or provide additional instructional intervention to keep them on track for at grade level reading outcomes.

2001

Assessment Committee

12

- Even the best intervention plans based on accurate and informative diagnostic assessment sometimes do not work as intended.
- Some estimates are that about 1 child in 3 does not respond to additional instructional intervention with the expected levels of achievement.
- An assessment system must include progress monitoring assessment to provide ongoing, formative evaluation of interventions so they can be modified early enough to achieve the major accountability outcomes for each and every child.
- Progress monitoring assessment should be closely aligned with intervention content, which should be closely aligned with accountability outcomes. The whole assessment, instruction, and intervention system should be aligned to the important beginning reading core areas.



Importance of Progress Monitoring

- When a hunter is lost in the woods,
 When a hikers are trying to find their way on a new trail,
 When a driver is looking for an address in an unfamiliar city,
 When a pilot is having difficulty finding the airport,
 When a skipper is trying to find the port in the fog,
 We have a technology to assist them in reaching their goal: *Global Positioning System or GPS* tells us,
- Where we are
 - Where we want to be
 - What course to follow
 - Our progress toward the goal.

2001

Assessment Committee

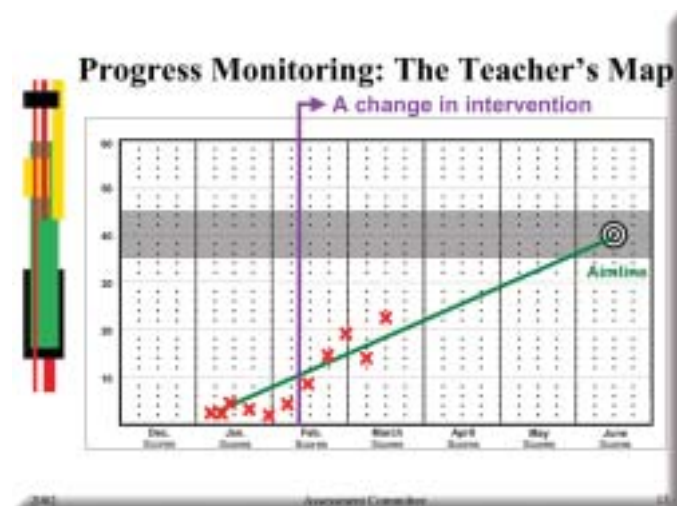
13

- The crucial role of monitoring progress toward an important goal is a very general principle for many endeavors. Hikers, drivers, pilots, and skippers all monitor where they are, where they want to be, what course they are following, and their progress toward the goal. They all modify their course if they see that their initial plan is not working in the way they anticipated.
- The Global Positioning System (GPS) is helpful technology to assist.

- For example, in the Northwest boating is an important recreation and livelihood. Whether you are on a whale watching tour or fishing, sometimes finding your way back to your port is easy. The sky is clear, the ocean blue, and you can clearly see your home port and the course you should follow to reach a safe harbor.
- But sometimes the fog rolls in and our journey to our goal becomes much more difficult and challenging. It is hard to tell where we are, where we want to be, what course to follow, and whether we are getting closer to safety or need to make a course adjustment.
- So we turn on the GPS and ask where we are. Of course, knowing where we are is only of limited help. The great philosopher Buckaroo Bonzai once commented, "No matter where you go, there you are!"
- We also need to know where the port, our safe harbor, is.
- We also need to know what course to follow to get there. The GPS can tell us to point the boat at 117 degrees and progress for 20 minutes at 10 knots to reach our goal.
- Now we have a good plan about how to get to our goal, our safe harbor, and avoid the rocks and cliffs on either side. But, sometimes our plans go awry....
- We also need to check up on our progress in time to make course corrections. If we are off course, the time to modify our plan is early, in time to still reach our safe harbor and not end up on the rocks.



- This is exactly the role of progress monitoring toward important reading outcomes.
- Progress monitoring provides a GPS system for educators.
- In this figure, the first 3 X's represent the child's initial level of skills with respect to the accountability goal for the grade.
- Across the bottom is month of the school year.
- The bulls eye represents the accountability outcome for the child's grade in school. If the child achieves that level of skill (or higher) then the child is on grade level and on track to meet standards for state accountability assessments in later grades.
- The green line represents the course the child will need to follow to achieve the desired outcome.
- The purple line represents a change or modification in the intervention plan. Sometimes the change may be a different intervention, sometimes the change may be additional time or practice opportunities within the same intervention. Sometimes you need a major course correction, sometimes a minor course correction will do.
- But, the *child must stay on course* to achieve the outcome.



Reading First Initiative: *At Grade Level*

States need to know, annually, whether they

- a) are making progress towards reducing the number of students who are reading below grade level,
- b) have significantly increased the number of students reading *at grade level* or above; and
- c) have significantly increased the percentages of students in ethnic, racial, and low-income populations who are reading *at grade level* or above.

- Key to the assessment system is a short, vivid indication of when students are on track for state accountability outcomes.
- At grade level is used here to convey that students either meet expectations on a state accountability assessment or are progressing at a rate that predicts they will meet expectations on a state accountability assessment administered in a later grade.

Defining *At Grade Level*

- **Rejected Definition:** Grade equivalents are rejected as a basis for determining *at grade level* for technical reasons.
- **Suggested Definition:** Students are *at grade level* if they meet expectations for reading proficiency on a state assessment or are predicted to meet expectations.
- **Secondary Definition:** *At grade level* is often used to refer to the *average* or *typical* reading skill for a grade level. A fundamental difficulty with average performance as a standard is that it is not possible to get most students *at grade level*.

- The words *at grade level* may lead some to think in terms of grade equivalents on achievement tests. However, age and grade equivalents suffer such severe limitations as a type of score that they have no redeeming social value and should not be used for any purpose.
- The most important aspect of performance *at grade level* is if the student meets expectations on a state accountability assessment administered at that grade level. Performance below expectations would be interpreted as *below grade level*.
- At grade levels where a state accountability assessment is not given, longitudinal evidence is needed to establish performance levels where students would be *predicted to meet* expectations on a later state accountability assessment.
- Referencing *at grade level* to state accountability assessments is preferred and will work well for most states. Some states may need an alternative approach.
- An alternative to state accountability assessment to establish *at grade level* is to use a norm-referenced test to identify average or typical performance as *at grade level*.

- States may employ a state accountability assessment, or they may specify explicit measurable standards for children's skills at each grade. Either may be used to determine that students are *at grade level*.



Suggested Ways to Show *At Grade Level*

- State Assessment:** Many states offer a state assessment at the end of third or fourth grade. Students are judged *proficient* or *not proficient* on the basis of the assessment. Students rated *proficient* or the equivalent on a state assessment are *at grade level*.
- State Standards:** Many states have explicit, measurable standards for performance at each grade level. Students meeting measurable state standards are *at grade level*.

2002

Assessment Committee

18

- Normative standing may be used to approximate meeting a state standard. In general, students performing at least at the 40th percentile are likely to meet state accountability standards.
- A fundamental problem with referencing *at grade level* to average or typical performance is that educational reform is likely to be a multi-year endeavor -- and it is likely to make a substantial difference in reading outcomes. If we as a nation are successful in improving the reading achievement of all of our children and especially those who would otherwise be poor readers, the norms will change. As tests are revised and renormed, average or typical performance will improve. Only in Lake Woebegone can we set a goal to have all children be at least average or typical readers.



Suggested Ways to Show *At Grade Level*

- Predicted to Meet State Standards:** In grades where a state assessment is not given, students are *at grade level* who are likely to meet the state standard in the next grade where a state assessment is given.
- Normative Standing:** A common use of *at grade level* is the level of performance that is typical for the grade. Students scoring at the 40th percentile or higher are *at grade level*.

2002

Assessment Committee

19

- Students who are *below grade level* will need additional instructional intervention to be *at grade level* on subsequent accountability assessments. Each accountability assessment serves a step toward the next accountability assessment.



Suggested Definition of *Needs Additional Intervention*

- Students who will need additional instructional intervention to achieve grade level outcomes:
 - Score somewhat but not severely below state standards.
 - May not meet state standards in third or fourth grade without additional intervention.
 - Score between the 20th and 39th percentile on an appropriate, nationally norm-referenced measure.

2002

Assessment Committee

20

Suggested Definition of *Needs Substantial Intervention*

- Students who will need substantial additional instructional intervention to achieve grade level outcomes:
 - Score *well below* state standards or expectations,
 - Are unlikely to meet state standards by third or fourth grade without *substantial* additional intervention,
 - Score below the 20th percentile on an appropriate, nationally norm-referenced measure.

- Students who are well below state standards or who score in the lowest 20 percent on a norm-referenced measure are likely to require substantial additional instructional intervention to achieve state standards in subsequent grades.
- The essential implication of performance below grade level is the provision of additional instructional intervention to reach grade level.

Assessment Menus for Reading First

- To support districts in selecting tests of the 5 important beginning reading core areas corresponding to the recommended assessment framework,
 - Criteria were developed to review tests,
 - Tests are being evaluated for use, based on the criteria, and
 - Menus of selected tests will be recommended
- A variety of recommended assessment *menus* will be developed. Each assessment menu will lay out a sequence of assessments that meet established criteria and that are logistically feasible.

- The Assessment Committee is developing a set of recommended assessment menus that a state or district can use to select assessments that will accomplish the purposes of outcome, screening, diagnostic, and progress monitoring assessment described here.

Process and Criteria for Selecting Reading Measures

The Reading First Assessment Committee has developed a set of processes and criteria to select, review, evaluate, and recommend reading assessment menus. The following steps are being followed:

1. Establish criteria to evaluate reading measures
2. Select reading measures for review
3. Describe logistical requirements of test use
4. Establish review and recommendation process

- The Reading First Assessment Committee is engaged in a systematic and careful process to create assessment menus.

- The technical adequacy of each assessment will be evaluated and reviewed prior to recommendation.

Step 1: Establish Criteria to Evaluate Reading Measures

The Committee's criteria to evaluate reading measures are based on the following questions:

- Does the test measure an important beginning reading skill?
- Does it provide sufficient information to assess whether the student is *at-grade level*?
- Is the test *reliable* (measures performance consistently) and *valid* (strongly relates to skill being measured)?
- Does the normative sample provide a meaningful comparison group for the students who will be assessed?

- Assessments must provide research-based evidence and must be aligned with the important skill areas of beginning reading.

Reading First Initiative: Rigorous Reading Assessment

Rigorous reading assessment means a reading assessment that—

- is valid, reliable, and grounded in scientifically based reading research;
- measures progress in phonemic awareness and phonics, vocabulary development, reading fluency, and reading comprehension; and
- identifies students who may be at risk for reading failure or who are having difficulty learning to read.

Reliable Assessment Is Essential

- Reliability of the assessment refers to the stability or consistency of test scores. To have confidence in assessment, we would expect a similar score if the students were tested:
 - On a different day.
 - By a different tester.
 - On a minimally different set of items.

Valid Assessment Is Essential

- Validity of assessment refers to evidence that the test measures what it is supposed to measure. A primary concern is that assessment measures the important beginning reading core areas:
 - Phonemic Awareness
 - Phonics
 - Fluency
 - Comprehension
 - Vocabulary
- Screening assessment must also provide evidence of predictive validity with respect to later reading outcomes.

2001

Assessment Committee

27

Step 2: Select Reading Measures for Review

The Reading First Assessment Committee selected reading measures for review based on the following criteria:

- a) Is the test frequently used in schools?
- b) Is the test frequently used in research/evaluation studies?
- c) Is the test prominent on lists developed by agencies and organizations?
- d) Is the test recommended by members of the Reading First Assessment Committee?

2001

Assessment Committee

28

Sample Tests to be Reviewed by the Committee

- Early Reading Diagnostic Assessment
- Woodcock Johnson Psycho-Educational Battery
- Woodcock Reading Mastery Test
- The Test of Word Reading Efficiency
- The Comprehensive Test of Phonological Processes, CTOPP
- The Test of Phonological Awareness
- The Phonological Awareness Test
- Gray Oral Reading Test-IV, GORT-4
- Texas Primary Reading Inventory

2001

Assessment Committee

29

Sample Tests to be Reviewed (continued)

- Lindamood Auditory Conceptualization Test
- Yopp-Singer Test of Phoneme Segmentation
- Qualitative Reading Inventory
- Iowa Test of Basic Skills
- Stanford Achievement Tests
- Terra Nova
- California Achievement Tests
- Auditory Analysis Test
- Roswell-Chall auditory Blending
- Dynamic Indicators of Basic Early Literacy Skills

2002

Assessment Committee

30

Sample Language Tests Reviewed

- Peabody Picture Vocabulary Test-III, PPVT-III
- Clinical Evaluation of Language Fundamentals-3rd, CELF-3
- Test of Language Development-Primary:3, TOLD-P:3
- Test of Word Knowledge

2002

Assessment Committee

31

Sample Spanish Tests for Review

- Test de Vocabulario en Imagenes Peabody, TVIP
- The Observation Survey (Spanish Equivalent)
- Developmental Reading Assessment (Spanish Equivalent)
- Woodcock-Munoz Language Survey
- Aprende: La Prueba de Logros en Espanol, Segunda Edicion
- Pre-Las 2000
- Spanish Reading Comprehension Test
- La Prueba de Realizacion, Segunda Edicion
- Spanish Assessment of Basic English, Second Edicion
- Tejas Lee (Texas Reads)

2002

Assessment Committee

32

Step 3: Describe Logistical Requirements of Test Use

The Committee established procedures to consider the following information about test use:

- Who administers the test? (e.g., teachers, aids, school psychologists)
- What are the administration formats and time requirements of the assessment? (e.g., group, individual)
- What does the test cost?
- What are the training requirements? (e.g., amount of training time, qualifications of testers)

- In addition to the technical requirements for tests used in the assessment system, they also must be logistically feasible for schools to use. They must be reasonably priced and not take too much time away from instruction.

Step 4: Establish Review and Recommendation Process

The Reading First Assessment Committee established the following process to review and select the reading tests:

- Frequently used and prominent measures will be reviewed using a standardized review form.
- A minimum of 2 qualified reviewers will analyze each reading measure.
- The Reading First Assessment Committee will review the findings and make the final decisions based on the extent to which the measures meet the evaluation criteria.

Now that You've Selected the Tests: Planning for Assessment

- Schedule the time to assess
- Train the testers or teachers
- Score tests
- Return information to teachers
- Help teachers to use the information to plan instruction and intervention
- Schedule regular sessions in which teachers discuss their students' scores and identify ways to incorporate the information into instruction
- Aggregate data across districts

- The first step in developing an assessment system for a district is to select the tests to be used. The assessment menus will help support that first step. Once the tests are selected, specific plans will need to be developed by each school district for implementing the assessment system.

- The remaining slides describe an optional small group activity that may be used by a local school district or school planning group to plan the assessment system.

Optional Small Group Activities

Action Plan for Implementing a District-Wide Early Assessment System

STEP I: Specify necessary steps to implement plan.

- Action: Meet with district- and school-level personnel to develop a plan to implement an early assessment system. Indicate in designated cells what is necessary to achieve each assessment purpose (i.e., screening, diagnostic, progress monitoring, outcome). Items in underline indicate necessary action.

2012

Assessment Committee

STEP II: Document degree of implementation.

- Action: Specify who is responsible and target completion date for each action below.

2012

Assessment Committee

Measure Selection

- Review list of measures and make final selection for each purpose.

Measure Acquisition

- Develop procedure to purchase and distribute measures to schools.

2012

Assessment Committee



Professional Development

- Review tester qualifications (per measure) and identify individuals.
- Provide training to ensure high quality test administration.
 - Specify who will administer measures, who will train data collectors, and procedures to ensure data are collected consistently.

2001

Assessment Committee

39



Data Collection Process and Schedule

- Specify when measures will be collected.
- Identify who will distribute materials.
- Specify where data will be collected.

2001

Assessment Committee

40



Scoring and Data Management

- Establish a secure and reliable method of scoring, entering, and managing data.
 - Specify who will:
 - ✓ score measures
 - ✓ enter data
 - ✓ manage database
 - ✓ cross-check data entry to ensure reliability
 - ✓ report to State DOE

2001

Assessment Committee

41



Information Reporting and Use

- Determine when & how information/results will be disseminated to teachers.
- Provide professional development on how to use data to inform instruction.
 - Schedule feedback and professional development sessions.

2002

Assessment Committee

42



Complaints You Might Hear:

- We're already trying to do too much.
- We don't have time to administer these assessments.
- Who's going to do all this?
- Who's going to pay for all this?
- All this testing isn't good for young children.
- These assessments are not authentic.
- What am I supposed to do with these results?
- I won't get the results back in time to do anything about it.
- Across assessments, aren't we combining apples and oranges?

2002

Assessment Committee

43




What Could Go Wrong? How to Avoid Pitfalls and Address Concerns

- Plan sufficient professional development regarding the importance of early intervention and the important beginning reading core areas (See the National Reading Panel Report).
- Plan adequate time and resources to administer assessments
- Plan sufficient personnel to administer assessments
- Plan for rapid turnaround of results and reports to inform instruction
- Plan to act on the data: procedures to review the comprehensive reading program and a system to provide additional instructional intervention

2002

Assessment Committee

44



Four Controversial Issues in Assessment

- Assertion #1: These tests don't measure "real" reading (i.e., deriving or constructing meaning from text).
- Assertion #2: Testing children in kindergarten is not developmentally appropriate.
- Assertion #3: "Weighing cows won't make 'em fatter" (i.e., assessing children, in and of itself, will not increase student learning).
- Assertion #4: Is this just one more thing to do? I don't have time for this.

ACTION PLAN FOR IMPLEMENTING A DISTRICT- WIDE EARLY ASSESSMENT SYSTEM

Kindergarten -- Action Plan for Implementing a Districtwide Early Assessment System

STEP I: *Specify necessary steps to implement plan.*

Action: Meet with district- and school-level personnel to develop a plan to implement an early assessment system. Indicate in designated cells what is necessary to achieve each assessment purpose (i.e., screening, diagnostic, progress monitoring, outcome). **Items in bold indicate necessary action.**

STEP II: *Document degree of implementation.*

Action: Specify who is responsible and target completion date for each action below.

<ul style="list-style-type: none"> • Measure Selection Review list of measures and make final selection for each purpose. • Measure Acquisition Develop procedure to purchase and distribute measures to schools. Specify final measures in the following cells and distribution process. 	<ul style="list-style-type: none"> • Professional Development Review tester qualifications (per measure) and identify individuals. Provide training to ensure high quality test administration. Specify who will administer measures, who will train data collectors, and procedures to ensure data are collected consistently. 	<ul style="list-style-type: none"> • Data Collection Process and Schedule Specify when measures will be collected. Identify who will distribute materials. Specify where data will be collected. 	<ul style="list-style-type: none"> • Scoring and Data Management Establish a secure and reliable method of scoring, entering, and managing data. Specify who will score measures, who will enter data, who will manage database, who will cross-check data entry to ensure reliability, who will report to State DOE. 	<ul style="list-style-type: none"> • Information Reporting and Use Determine when & how information/results will be disseminated to teachers. Provide professional development on how to use data to inform instruction. Schedule feedback and professional development sessions.
---	---	---	---	---

Purpose of Measure: Screening

			Reporting to State Department not required	
--	--	--	--	--

Purpose of Measure: Diagnostic

			Reporting to State Department not required	
--	--	--	--	--

Purpose of Measure: Progress Monitoring

			Reporting to State Department not required	
--	--	--	--	--

Purpose of Measure: Achievement Outcomes

--	--	--	--	--

Grade 1 -- Action Plan for Implementing a Districtwide Early Assessment System

STEP I: *Specify necessary steps to implement plan.*

Action: Meet with district- and school-level personnel to develop a plan to implement an early assessment system. Indicate in designated cells what is necessary to achieve each assessment purpose (i.e., screening, diagnostic, progress monitoring, outcome). **Items in bold indicate necessary action.**

STEP II: *Document degree of implementation.*

Action: Specify who is responsible and target completion date for each action below.

<ul style="list-style-type: none"> • Measure Selection Review list of measures and make final selection for each purpose. • Measure Acquisition Develop procedure to purchase and distribute measures to schools. Specify final measures in the following cells and distribution process. 	<ul style="list-style-type: none"> • Professional Development Review tester qualifications (per measure) and identify individuals. Provide training to ensure high quality test administration. Specify who will administer measures, who will train data collectors, and procedures to ensure data are collected consistently. 	<ul style="list-style-type: none"> • Data Collection Process and Schedule Specify when measures will be collected. Identify who will distribute materials. Specify where data will be collected. 	<ul style="list-style-type: none"> • Scoring and Data Management Establish a secure and reliable method of scoring, entering, and managing data. Specify who will score measures, who will enter data, who will manage database, who will cross-check data entry to ensure reliability, who will report to State DOE. 	<ul style="list-style-type: none"> • Information Reporting and Use Determine when & how information/results will be disseminated to teachers. Provide professional development on how to use data to inform instruction. Schedule feedback and professional development sessions.
---	---	---	---	---

Purpose of Measure: Screening

			Reporting to State Department not required	
--	--	--	--	--

Purpose of Measure: Diagnostic

			Reporting to State Department not required	
--	--	--	--	--

Purpose of Measure: Progress Monitoring

			Reporting to State Department not required	
--	--	--	--	--

Purpose of Measure: Achievement Outcomes

--	--	--	--	--

Grade 2 -- Action Plan for Implementing a Districtwide Early Assessment System

STEP I: *Specify necessary steps to implement plan.*

Action: Meet with district- and school-level personnel to develop a plan to implement an early assessment system. Indicate in designated cells what is necessary to achieve each assessment purpose (i.e., screening, diagnostic, progress monitoring, outcome). **Items in bold indicate necessary action.**

STEP II: *Document degree of implementation.*

Action: Specify who is responsible and target completion date for each action below.

<ul style="list-style-type: none"> • Measure Selection Review list of measures and make final selection for each purpose. • Measure Acquisition Develop procedure to purchase and distribute measures to schools. Specify final measures in the following cells and distribution process. 	<ul style="list-style-type: none"> • Professional Development Review tester qualifications (per measure) and identify individuals. Provide training to ensure high quality test administration. Specify who will administer measures, who will train data collectors, and procedures to ensure data are collected consistently. 	<ul style="list-style-type: none"> • Data Collection Process and Schedule Specify when measures will be collected. Identify who will distribute materials. Specify where data will be collected. 	<ul style="list-style-type: none"> • Scoring and Data Management Establish a secure and reliable method of scoring, entering, and managing data. Specify who will score measures, who will enter data, who will manage database, who will cross-check data entry to ensure reliability, who will report to State DOE. 	<ul style="list-style-type: none"> • Information Reporting and Use Determine when & how information/results will be disseminated to teachers. Provide professional development on how to use data to inform instruction. Schedule feedback and professional development sessions.
---	---	---	---	---

Purpose of Measure: Screening

			Reporting to State Department not required	
--	--	--	--	--

Purpose of Measure: Diagnostic

			Reporting to State Department not required	
--	--	--	--	--

Purpose of Measure: Progress Monitoring

			Reporting to State Department not required	
--	--	--	--	--

Purpose of Measure: Achievement Outcomes

--	--	--	--	--

Grade 3 -- Action Plan for Implementing a Districtwide Early Assessment System

STEP I: *Specify necessary steps to implement plan.*

Action: Meet with district- and school-level personnel to develop a plan to implement an early assessment system. Indicate in designated cells what is necessary to achieve each assessment purpose (i.e., screening, diagnostic, progress monitoring, outcome). **Items in bold indicate necessary action.**

STEP II: *Document degree of implementation.*

Action: Specify who is responsible and target completion date for each action below.

<ul style="list-style-type: none"> • Measure Selection Review list of measures and make final selection for each purpose. • Measure Acquisition Develop procedure to purchase and distribute measures to schools. Specify final measures in the following cells and distribution process. 	<ul style="list-style-type: none"> • Professional Development Review tester qualifications (per measure) and identify individuals. Provide training to ensure high quality test administration. Specify who will administer measures, who will train data collectors, and procedures to ensure data are collected consistently. 	<ul style="list-style-type: none"> • Data Collection Process and Schedule Specify when measures will be collected. Identify who will distribute materials. Specify where data will be collected. 	<ul style="list-style-type: none"> • Scoring and Data Management Establish a secure and reliable method of scoring, entering, and managing data. Specify who will score measures, who will enter data, who will manage database, who will cross-check data entry to ensure reliability, who will report to State DOE. 	<ul style="list-style-type: none"> • Information Reporting and Use Determine when & how information/results will be disseminated to teachers. Provide professional development on how to use data to inform instruction. Schedule feedback and professional development sessions.
---	---	---	---	---

Purpose of Measure: Screening

			Reporting to State Department not required	
--	--	--	--	--

Purpose of Measure: Diagnostic

			Reporting to State Department not required	
--	--	--	--	--

Purpose of Measure: Progress Monitoring

			Reporting to State Department not required	
--	--	--	--	--

Purpose of Measure: Achievement Outcomes

--	--	--	--	--

THE IMPORTANCE AND DECISION-MAKING
UTILITY OF A CONTINUUM OF FLUENCY-BASED
INDICATORS OF FOUNDATIONAL READING
SKILLS FOR THIRD GRADE HIGH STAKES
OUTCOMES: AN INTRODUCTORY PAPER

The Importance and Decision-Making Utility of A Continuum of Fluency-Based Indicators of Foundational Reading Skills for Third-Grade High-Stakes Outcomes

Roland H. Good III

Deborah C. Simmons

Edward J. Kame'enui

University of Oregon

Abstract

Educational accountability and its counterpart, high-stakes assessment, are at the forefront of the educational agenda in this era of standards-based reform. In this article, we examined assessment and accountability in the context of a prevention-oriented assessment and intervention system designed to assess early reading progress formatively. Specifically, we explored the utility of a continuum of fluency-based indicators of foundational early literacy skills to predict reading outcomes, to inform educational decisions, and to change reading outcomes for students at risk of reading difficulty. First, we addressed the accountability era, discussed the promise of prevention-oriented assessment, and outlined a continuum of fluency-based indicators of foundational reading skills using Dynamic Indicators of Basic Early Literacy Skills (DIBELS) and Curriculum-Based Measurement Oral Reading Fluency (CBM ORF). Next, we described a series of linked, short-term, longitudinal studies of four cohorts examining the utility and predictive validity of the measures from kindergarten through third grade with the Oregon Statewide Assessment-Reading/Literature as a high-stakes, reading outcome. Using direct measures of key foundational skills, predictive validities ranged from .34 to .82. The utility of the fluency-based benchmark goals was supported with the finding that 96% of children who met the third-grade oral reading fluency benchmark goal met or exceeded expectations on the Oregon Statewide Assessment, a high-stakes outcome measure. We illustrated the utility of the measures for evaluating instruction, modifying the instructional system, and targeting children who need additional instructional support to achieve benchmark goals. Finally, we discussed the instructional and policy implications of our findings and their utility in an active educational accountability environment.

The Importance and Decision-Making Utility of A Continuum of Fluency-Based Indicators of Foundational Reading Skills for Grade Three High-Stakes Outcomes

Across the nation, there is growing awareness of the dividends of early reading success and the stark consequences of early reading failure. Though the reading levels of students in the United States remained relatively stable over the past two decades (National Center for Education Statistics, 1998), unlike previous generations these reading proficiency levels no longer satisfy today's societal requirements and aggressive economic environment. The demands of the knowledge-based, 21st-century workplace (Drucker, 1993; Murnane & Levy, 1996) have raised the literacy bar for America's students, and schools must now respond in kind to heightened expectations. One of the most promising strategies to address this monumental goal is to prevent reading difficulties and to ensure that all children are readers early in their educational careers (National Research Council, 1998).

Though the goal of children reading by Grade 3 is not altogether new, the proposed policies and practices to achieve this goal are. The past ten years ushered into education an unfamiliar vocabulary and unique set of policies and practices designed to address the problem of low achievement in America's schools. Terms such as standards-based reform, accountability, and high-stakes assessment (Carnine, 2000; Thurlow & Thompson, 1999) were relatively disassociated with education a decade ago, but are now part of the educational rhetoric. Though standards-based reform has multiple dimensions, the component that is most prominent and polarizing is the process of "using assessments for accountability purposes" (Thurlow & Thompson, 1999, p. 3).

The high-stakes accountability movement calls for an assessment system that produces trustworthy and reliable results that are instructionally relevant and capable of forecasting educational change that positively impacts and sustains student learning (Carnine, 2000; Elmore, 1996; Linn, 2000). Typically, the first high-stakes assessment is administered in Grade 3. During the primary grades, an accountable assessment system would document whether students are learning "enough" (Carnine, 1997) before Grade 3 and before reading problems become too great and intractable. Such a system would allow reasonable and reliable predictions of whether children who perform well on one measure or set of measures in one year are likely to perform at designated benchmark levels in subsequent years.

In this article, we examine assessment and accountability in the context of prevention. First, we examine the accountability era, discuss the promise of a prevention-oriented assessment and intervention system, and propose a measurement model based on a continuum of fluency-based indicators of foundational reading skills. Next, we describe a series of linked, short-term longitudinal correlational and conditional probability analyses involving four cohorts of students enrolled in kindergarten through Grade 3. We examine student performance on early, fluency-based reading indicators and examine their utility in predicting reading success or failure on Grade 3 high-stakes reading achievement tests. Finally, we discuss the instructional and policy implications of our findings and their potential utility in an active educational accountability environment.

The Accountability Era and The Attributes of a Prevention-Oriented Assessment and Intervention System

Educational accountability and its counterpart, high-stakes assessment, are at the forefront of the educational agenda. For most states, the primary tool to evaluate students' knowledge and understanding of content standards is the standardized achievement test. Bond, Roeber, and Connealy (1998) reported that 31 states use normative-referenced tests, 33 use criterion-referenced measures, and 19 use both forms of standardized testing to assess student knowledge and understanding of state content standards. Commercial, standardized achievement tests, by design, are intended to provide "a level playing field" for comparing children on the same content and for determining proficiency in a given content or skill area (Green & Sireci, 1999). The tenets of fairness and content comparability are laudable and defensible, psychometrically. Nevertheless, traditionally administered commercial, standardized achievement tests have serious limitations in a high-stakes assessment system. Generally, the commercial, standardized reading achievement tests used in high-stakes assessments are time-consuming, expensive to administer, administered infrequently, and of limited instructional utility (Fuchs & Fuchs, 1999; Kame'enui & Simmons, 1990).

For the purposes of gauging district or school-wide progress and global levels of performance, large-scale, traditional assessments may serve an important function. However, for the purpose of informing instruction in time-efficient, instructionally relevant ways capable of altering students' rates and levels of learning on critical indicators of reading, commercial standardized measures are severely limited, if not inappropriate (e.g., Shephard, 2000). In his review of assessment and accountability over the past 50 years, Linn (2000) lamented that he could not conclude that the use of tests for student and school accountability has produced dramatic improvements in our education system or outcomes. He did conclude, however, that the "instruments and technology have not been up to the demands that have been placed on them by high-stakes accountability" (p. 14). In the following section, we outline the dimensions of a prevention-oriented, school-based assessment and intervention system designed to complement existing high-stakes assessment systems and pre-empt early reading difficulty from becoming established, inadequate reading achievement.

Assessment in a Prevention-Oriented Framework: Measuring What's Important

Though this study focuses on assessment, the broader focus is on the role of assessment in a comprehensive, integrated educational system. States design and sanction standards and the tests used to assess proficiency on those standards. Schools assume the fundamental responsibility for ensuring that all children read by Grade 3. States determine the level of proficiency required of students to clear the grade-level learning hurdle. Schools are directly accountable for all children being able to read by the end of Grade 3. In a prevention-oriented system, schools have the responsibility to design and use assessment and intervention that adheres to the following principles:

- (a) Intervene early and strategically during critical windows of reading development;
- (b) Develop and promote a comprehensive system of instruction based on a research-based core curriculum and enhancement programs;
- (c) Use and rely on formative, dynamic indicators of student performance to identify need, allocate resources, and design and modify instruction;
- (d) Address reading failure and reading success from a schoolwide systemic perspective.

Signature attributes of a prevention-oriented, school-based assessment and intervention system (Simmons, Kame'enui, Good III, Harn, Cole, & Braun, 2000) are the ability to predict reading success and difficulty early and to inform instruction responsively. An assessment system must be in place that signals reading difficulty early and prevents early reading risk from becoming entrenched reading failure (National Research Council, 1998; Torgesen, 1998). One of the most replicated and disturbing conclusions from studies of reading is that students with poor reading skills initially are likely to have poor reading skills later (e.g., Juel, 1988; Shaywitz, Escobar, Shaywitz, Fletcher, & Makuch, 1992). Differences in developmental reading trajectories can be explained, in part, by a predictable and consequential series of reading-related activities that begin with difficulty in foundational skills, progress to fewer encounters and exposure to print, and culminate in lowered motivation and desire to read (Stanovich, 1986; Stanovich, 2000). Low initial skills and low learning trajectories make catching up all but impossible for many readers at risk for reading difficulties. In an era of high-stakes outcomes, the message is clear: We must have a reliable prevention-oriented, school-based assessment and intervention system to prevent early reading difficulty from forecasting enduring and progressively debilitating reading failure. That assessment system must be dynamic in the sense that it is able to measure and track changes in student performance over time.

Assessment for educational prevention and accountability requires more than just a new test; it requires a different conceptual approach. In the primary grades, such an assessment system in schools at minimum must reliably (a) document and account for growth on a continuum of foundational reading skills, (b) predict success or failure on criterion measures of performance (i.e., high-stakes tests), and (c) provide an instructional goal that if met will prevent reading failure and promote reading success. Such an assessment system is based on the assumption that the measures not only document whether students are learning but whether they are learning enough prerequisite, foundational skills in a timely manner to attain benchmark levels on high-stakes tests. Moreover, the utility and validity of the assessment system is grounded in two fundamental features: (a) identifying the foundational skills of beginning reading, and (b) evaluating growth of foundational skills efficiently and reliably.

Measuring What's Important: The Foundational Skills of Beginning Reading

It is generally recognized that reading is developmental and acquired over time. Multiple models of reading articulate the stages of reading development (e.g., Chall, 1983; Ehri & McCormick, 1998). Despite modest differences in theory and nomenclature, there is considerable congruity among models regarding the critical dimensions of reading development. Converging and convincing evidence substantiates that reading competence is causally influenced by proficiency on foundational skills in beginning reading (National Reading Panel, 2000; National Research Council, 1998). Among the commonly recognized and empirically validated foundational skills are skills we refer to as “big ideas” in beginning reading. Big ideas are skills and strategies that are prerequisite and fundamental to later success in a content area or domain. They are skills that differentiate successful from less successful readers and most important are amenable to change through instruction (Kame'enui & Carnine, 1998; Simmons & Kame'enui, 1998). In the area of beginning reading, selected foundational skills include: (a) phonological awareness or the ability to hear and manipulate the sound structure of language, (b) alphabetic understanding or the mapping of print to speech and the phonological recoding of letter strings into corresponding sounds and blending stored sounds into words, and (c) accuracy and fluency

with connected text or the facile and seemingly effortless recognition of words in connected text (Adams, 1990; National Reading Panel, 2000; National Research Council, 1996; Simmons & Kame'enui, 1998).

While these three foundational skills and processes are by no means exhaustive of beginning reading and early literacy, they represent valid indicator skills along a continuum in which overlapping stages progress in complexity toward an ultimate goal of reading and constructing meaning from a variety of texts by the end of Grade 3. In a prevention-oriented assessment and intervention system, these foundational skills can be assessed early (e.g., fall of kindergarten) and monitored over time as the foci of instruction change and children's reading skills develop more expansively and comprehensively.

Measuring growth of foundational skills

The concept of growth is fundamental to any comprehensive discussion of assessment (Francis, Shaywitz, Stuebing, Shaywitz, & Fletcher, 1994). Measuring early reading growth in a prevention-oriented assessment and intervention system requires measures and methodology that (a) first and foremost measure growth reliably and validly, (b) specify criterion-levels of performance for a single measure, (c) assess performance on a continuum of linked measures that relate to one another, and (d) reliably document a child's progression toward meaningful outcomes. The goal for prevention-oriented assessment is to equip schools with a measurement system that reliably predicts performance on critical outcomes early and in ways that are relevant to instruction. Core to this system are instruments that are capable of measuring beginning reading growth functionally and frequently in the complex host environments of schools (O'Connor, 2000; Simmons et al., 2000; Torgesen, 2000) where time is finite and resources are fixed. We propose that reading fluency-based indicators readily lend themselves to these purposes and conditions.

The foundation of the prevention-oriented assessment and intervention system was laid more than 20 years ago with the work of Stan Deno and colleagues (see Fuchs, Fuchs, Hosp, & Jenkins, this issue). This measurement methodology, known as curriculum-based measurement (CBM), is perhaps best known in the particular application of CBM oral reading fluency (ORF). CBM ORF was developed as a method to measure increased reading proficiency based on scoring frequent, short-duration performance samples obtained by having students read aloud from text passages of equivalent difficulty (Deno, Mirkin, & Chiang, 1982). The procedures used to obtain these repeated samples of reading performance are an example of general outcome measurement (Fuchs & Deno, 1991) in which the number of words read correctly from passages in one minute is representative of the curriculum and deemed to serve as broad indicators of reading competence (see Fuchs et al., this issue). The content, criterion, and construct validities of CBM as well as alternate-form and test-retest reliabilities are well documented and substantiated (Fuchs & Deno, 1991; Markell & Deno, 1997). Original purposes of CBM were to serve as an objective tool for identifying students (a) who were discrepant from classroom peers, and (b) in need of diagnostic assessment (Fuchs & Deno, 1991). Furthermore, CBM has been used to evaluate students' rate of progress and to evaluate the efficacy of instruction. The advantages of CBM ORF in a prevention-oriented model are logically intuitive and empirically validated. The limitation of this measure, however, is that most children do not have sufficient proficiency with connected text to measure reading validly until mid-to-late first grade. In a prevention-oriented assessment and intervention system, the need for measures that document

growth on other critical indicators in the foundational skills of reading acquisition is essential. Central to this methodology is the role of fluency.

The Role of Fluency in Early Reading Assessment

In reading, fluency is most commonly construed as oral reading fluency in connected text. The National Reading Panel (NRP) (2000) defined fluency as “the ability to read a text quickly, accurately, and with proper expression” (chap. 3, p. 5), and through a quantitative meta-analysis of 77 research studies corroborated fluency’s importance in overall reading competence. Fluency is an important focus of instruction that encompasses but extends beyond accurate word recognition and is a causal determinant of higher-order skills such as reading comprehension (NRP, 2000). Beyond defining and documenting the importance of fluency to reading, the NRP expertly chronicled the evolution of fluency and automaticity outlining critical dimensions and contributions to reading.

Automaticity or fluency in cognitive processes such as reading involves more than the seemingly quick and effortless access to information. Automaticity involves the “processing of information that ordinarily requires long periods of training before the behavior can be executed with little effort or attention” (NRP, 2000, chap. 3, p. 7). Additional properties the NRP derived from classical studies of cognitive and experimental psychology note that automaticity (a) happens gradually (Shiffrin & Schneider, 1977), (b) occurs without immediate intention (Posner & Snyder, 1975), (c) allows for parallel processing of other information (Ackerman, 1987), and (d) occurs along a continuum rather than a dichotomy (Logan, 1997a). Whether one ascribes to the resource-capacity theory (LaBerge & Samuels, 1974), the two-process theory of expectancy (Posner & Snyder, 1975), or the information encapsulation theory of automaticity (Logan, 1988; 1997b), at a rudimentary level the common denominator among the three theoretical bases is that speed of processing is a proxy for level of learning. As skills are learned, the time required to produce the response can be used as an indicator of proficiency. Analysis and comparison of the differing theories of automaticity is beyond the scope of this paper; nevertheless, the processes that differentiate learners' response rates is critically important for future instruction.

In our application, fluency is not limited to reading connected text quickly, accurately, or with proper expression. Instead, it incorporates the development of the prerequisite and foundation skills of beginning reading such as phonemic awareness, alphabetic understanding, and phonological recoding and the need for a high criterion-level of proficiency of each. Moreover, it is predicated on the proposition that fluent performance of complex skills and higher-level processes (e.g., word recognition and reading comprehension) requires fluency in the component skills and lower-level processes. Several recent fluency studies have targeted word recognition and demonstrated gains in connected-text fluency and comprehension (Levy, Abello, & Lysynchuk, 1997; Tan & Nicholson, 1997 cited in Wolf, Bowers, & Biddle, 2000). Wolf, Bowers, and Biddle (2000) and others (e.g., Torgesen, 1998) noted, however, that interventions that address automaticity in the foundational skills that service word and text-level processing have received little sustained attention.

The premise of assessment examined in this study is that fluency as represented by accuracy and rate pervades all levels of processing involved in reading (Logan, 1997b) and that fluency on early foundation skills can be used to predict proficiency on subsequent skills in reading. To evaluate the role and relation of fluency in the development of foundation skills in beginning reading and Grade 3 high-stakes reading achievement, we employed a continuum of fluency-based measures developed and validated for use with children in kindergarten and early

first grade called the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) (Kaminski & Good, 1996). We complemented DIBELS with CBM ORF in Grades 1-3. DIBELS measures were designed to assess students' early literacy skills dynamically as they change over time. As such, these measures are sensitive to student growth, easy and efficient to administer (e.g., each measure is a one-minute, fluency-based measure), capable of repeated and frequent administration (e.g., the Phonemic Segmentation Fluency measure has 25 alternate forms of equivalent difficulty), and cost effective (Good, Simmons, & Smith, 1998). DIBELS are not designed to serve as a comprehensive or diagnostic reading assessment tool. Rather, they are intended to "provide a fast and efficient indication of the academic well-being of students with respect to important early literacy skills" (Good, Simmons, & Smith, p. 748) and represent an efficient and parsimonious approach to early literacy assessment.

A Preventive Measurement Model: Conceptual, Procedural, and Developmental Dimensions

Few would argue with the concept of prevention and the need for formative assessment to inform instruction. In the following figure, we make concrete the conceptual and procedural dimensions of such a measurement model and outline a developmental timeline for the acquisition of crucial reading skills (See Figure 1). The top level of ellipses summarizes the conceptual dimensions of reading acquisition that include three "big ideas" of beginning reading: (a) phonological awareness, (b) alphabetic principle, (c) and accuracy and fluency with connected text. These big ideas provide a foundation for meeting expectations on high-stakes outcome measures of reading proficiency. This model is not intended to capture all of the complexities and nuances of reading acquisition, but to represent key skills within the instructional domain that are necessary but not sufficient for successful reading.

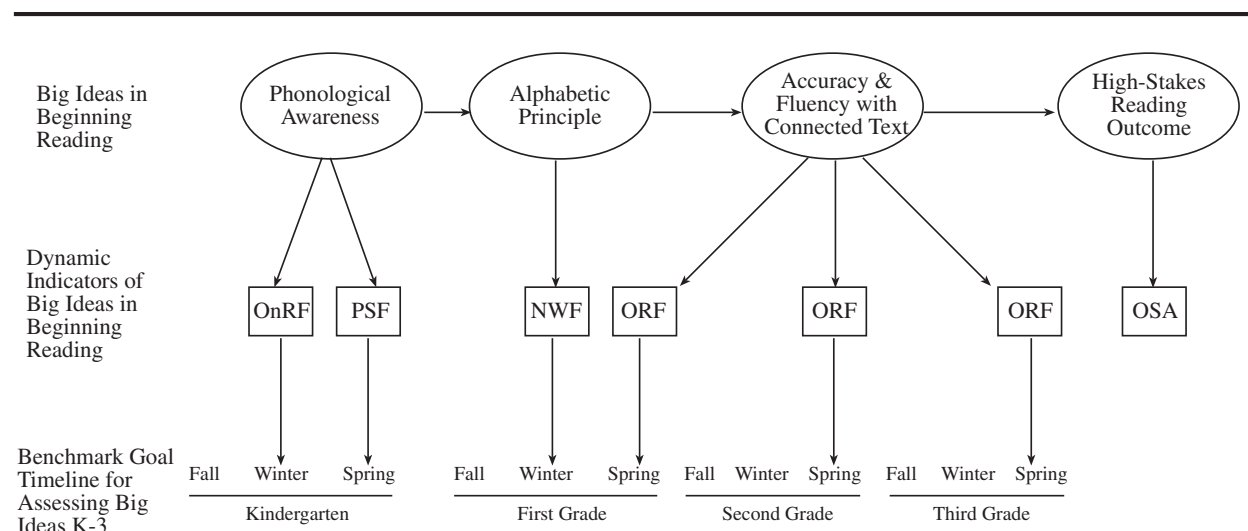


Figure 1. Conceptual and procedural dimensions and timeline for acquisition of reading and early literacy skills.

The second level of rectangles in Figure 1 summarizes the procedural dimensions and specifically the fluency-based measures, which provide an efficient indication of the acquisition of the big ideas of early reading. The third level of the model provides a timeline for the

acquisition of reading skills necessary to meet expectations on high-stakes measures of reading outcomes. By combining a level of skill and a timeline for acquisition, benchmark goals can be established. Thus, Onset Recognition Fluency (OnRF) provides an indicator of the child's knowledge and awareness of initial sounds in words, an aspect of phonological awareness desired by winter of kindergarten if the child is to be on track for reading outcomes. Phoneme Segmentation Fluency (PSF) provides an indicator of phonological awareness skills necessary by spring of kindergarten. By winter of first grade, students should display alphabetic principle skills on Nonsense Word Fluency (NWF), and by spring of first grade, they should reach target levels of Oral Reading Fluency (ORF), a measure of accuracy and fluency with connected text. By spring of second grade and spring of third grade, adequate progress on measures of ORF is necessary to be on track for high-stakes reading outcomes. The model is designed to make explicit a set of parsimonious linkages between earlier and later skills at different points in time.

The timing of these benchmark goals specifies when target levels of phonological awareness, alphabetic principle, and accuracy and fluency with connected text skills should be attained. Instruction and curriculum should be emphasizing those skills prior to the benchmark goal timing. In addition, assessment of target skills also should occur earlier than the outcome time in order to allocate resources and monitor progress toward the benchmark goal.

Initial Establishment of Benchmark Goals

The establishment of benchmark goals is a challenging, but important task. For teachers knowing which skill areas are crucial for early literacy is an important first step, but of likewise importance is knowing how proficient children are in these critical skills. To understand how much of the skill is desired to provide a sound foundation for later literacy skill acquisition we quantify early literacy proficiency using benchmarks. An effective benchmark goal should be specific, measurable, ambitious, and target a critical indicator of student performance (Fuchs et al., 1993). Equally important, a benchmark goal should be linked to or anchored by a socially meaningful and important outcome. Ideally, establishment of a benchmark goal integrates statistical, psychometric, and socio-political considerations in an overall judgment.

The approach to benchmark goal setting followed in this program of research has been to first set an initial estimate of a goal based on the best available empirical evidence, theoretical rationale, and judgment of social value. Then, the utility of the initially specified benchmark goal is examined in different contexts, with different samples of children, and at different times. Based upon the utility of the goal and responses of users, the goal may be modified and re-examined. This study falls in the middle of a program of research on goal approximation and evaluation (Good et al., in preparation; Simmons, Kame'enui, & Good, 1998). Initial establishment of benchmark goals followed different procedures for (a) spring-of-first grade benchmark goal, (b) DIBELS benchmark goals, and (c) spring-of-second and third-grade CBM ORF benchmark goals.

Spring-of-first grade reading benchmark goal. The anchor for the system of benchmark goals represented by the prevention-oriented assessment and intervention system described here was all first-grade students reading at or above 40 words read correct per minute on grade-level material using CBM oral reading fluency procedures at the end of the year. It is important to note that this goal is not the goal for the average student in first grade – it is the goal for all students in first grade, including the lowest readers. If all children are to be readers by third grade (National Research Council, 1998), then all children must make satisfactory reading progress in first grade.

Support for the benchmark goal of 40 or more on CBM ORF in spring of first grade for all students derives from empirical, theoretical, and social-validation sources. First, 40 or more on CBM ORF is associated with a trajectory of reading progress with an adequate slope of progress. Good, Simmons, and Smith (1998) contrasted the trajectories of progress for students in the middle 10% of a district at the beginning of second grade with the trajectories of progress for students in the lowest 10% at the beginning of second grade. Students in the middle 10% displayed trajectories of progress with positive slope and consistently had beginning second grade skills of 40 words correct per minute or higher on CBM ORF. Students entering second grade with CBM ORF scores approaching 10 or lower displayed substantially lower or zero slopes of progress, and fell increasingly further behind their regularly achieving peers.

A second criterion of an effective goal is rigor or ambitiousness. A goal should represent a reasonable yet rigorous target. For all first graders, 40 or more correct words per minute on CBM ORF is an ambitious goal. In examinations of district performance on CBM ORF, few districts have attained 100 percent of their students with skills above 40 at the end of first grade or beginning of second grade (Fuchs et al., 1993; Hasbrouck & Tindal, 1992). Nonetheless, a goal of 40 or more on CBM ORF for all or almost all students appears attainable. For example, Lyon 1997), in summarizing 15 years of NICHHD research reported, “we have learned that for 85 to 90 percent of poor readers, prevention and early intervention programs that combine instruction in phoneme awareness, phonics, spelling, reading fluency, and reading comprehension strategies provided by well-trained teachers can increase reading skills to average reading levels” (Lyon, 1997). Finally, 40 or more on CBM ORF appears to be socially meaningful and important.

Establishing DIBELS benchmark goals. Initial establishment of benchmark goals for the DIBELS measures was conducted through the Early Childhood Research Institute (ECRI) on Measuring Growth and Development at the University of Oregon (Good, Kaminski, M. R. Shinn, Bratten, M. M. Shinn, & Laimon, in preparation). The development of the benchmark goal for DIBELS PSF is illustrative of the process followed for all early literacy indicators. As a part of the ECRI longitudinal research on the DIBELS measures, all kindergarten children ($n = 78$) in an elementary school were assessed with DIBELS PSF in spring of kindergarten. One year later, in the spring of first grade, all first-grade children were assessed on the CBM ORF measure. Due to high child mobility in the school, 56 children had both kindergarten and first-grade assessments. Kindergarten PSF was significantly correlated with first-grade CBM ORF ($r = .62$) and the scatterplot illustrating the relationship is provided in Figure 2.

The top horizontal line on Figure 2 at a CBM ORF score of 40 corresponds to a level of reading skills judged to be an appropriate and desired outcome for first-grade readers. Students scoring at or above the line at 40 would be judged to have attained an appropriate and desired level of reading skills at the end of first grade. This judgment represents a key assumption upon which the establishment of early literacy benchmarks rests. The lower horizontal line at a CBM ORF score of 10 represents a problematic reading outcome. Students reading 10 or fewer words on a grade-level reading passage in a minute are struggling and experiencing significant reading difficulty. Students scoring between 10 and 40 on CBM ORF have emerging reading skills, but have not attained goal levels of reading skills for first grade.

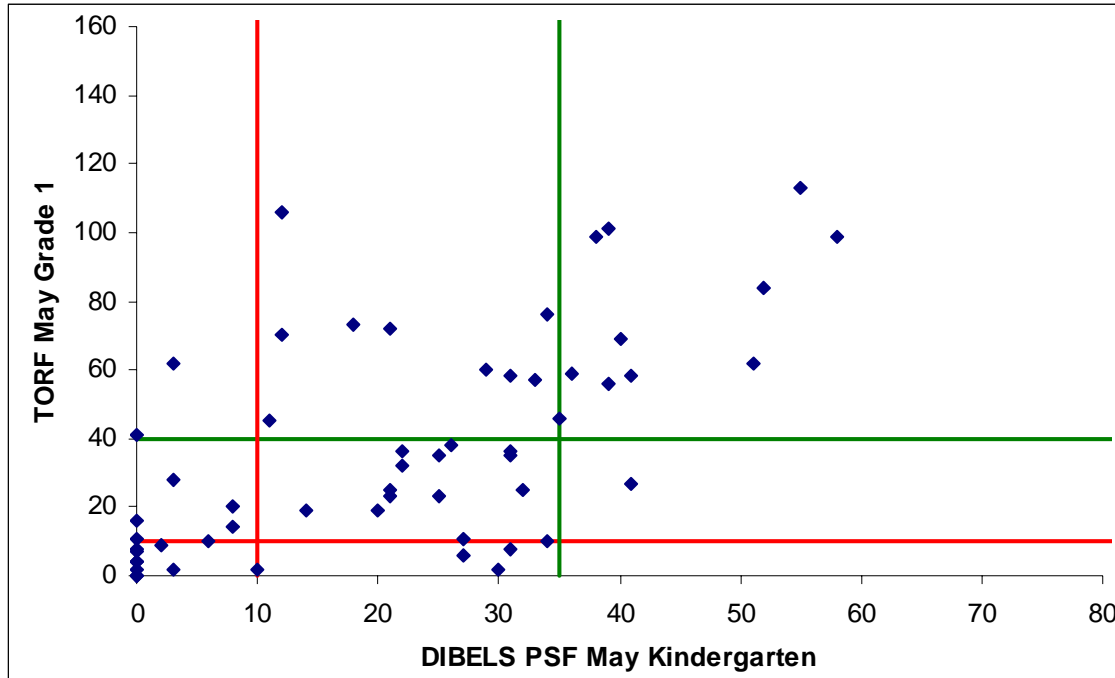


Figure 2. Initial establishment of benchmark goals based on the relation between spring of kindergarten Dynamic Indicators of Basic Early Literacy Skills Phoneme Segmentation Fluency and spring of first grade Curriculum-Based Measurement Oral Reading Fluency. TORF = Test of Reading Fluency.

From an examination of Figure 2, three levels of phonological awareness skills on the DIBELS PSF measure in spring of kindergarten were identified. The first group of students scored 35 or better on DIBELS PSF, and most of those students attained desired first-grade reading outcomes. Of the 12 students who scored 35 or above on DIBELS PSF in spring of kindergarten, 11 students (92%) read 40 or more words on CBM ORF in spring of first grade. The second group of students scored between 10 and 35 on DIBELS PSF in spring of kindergarten, and a clear prediction of reading outcomes was not possible. Some (35%) of the students scoring between 10 and 35 attained desired reading outcomes in spring of first grade, others experienced serious reading difficulty at the end of first grade. A third group of students received scores of 10 or lower on DIBELS PSF in spring of kindergarten and were clearly at risk for poor reading outcomes in first grade. Of the 18 students who scored 10 or lower, only 2 students (11%) attained desired reading outcomes at the end of first grade.

Using this procedure, 35 correct phonemes per minute on DIBELS PSF was established as an initial benchmark goal for spring of kindergarten. A kindergarten teacher who teaches phonological awareness skills well enough so that his or her students score 35 or better on DIBELS PSF in spring of kindergarten can be confident that his or her students are making adequate progress toward reading outcomes. It also appears that students scoring 10 or below will likely need intensive instructional support if they are going to attain desired reading outcomes by the end of first grade.

Based on similar analyses and logic, benchmark goals and timelines for a trajectory of desired progress toward high-stakes reading outcomes through spring of first grade were established for OnRF in winter of kindergarten and NWF in winter of first grade (Good et al., in preparation). These initial benchmark goals are summarized in Table 1. All of these early literacy benchmarks rely in some way upon the judgment that 40 or more on CBM ORF using grade-level material in spring of first grade represents a desired and appropriate level of reading

competence. Each benchmark represents a level of skill with respect to a big idea of early literacy where the student is likely to attain desired first-grade reading outcomes.

Establishing second- and third-grade CBM ORF benchmark goals. Benchmark goals for the end of second grade and the end of third grade build upon and extend the work of Hasbrouck and Tindal (1992). They found that, across multiple sites, the 50th percentile of correct words read per minute on grade-level passages in spring of first grade was 94, and the 50th percentile was 114 in spring of third grade. A problem with using 94 and 114 as goals is that they are based on normative expectations of performance that may not necessarily correspond to desired and appropriate outcomes. A level of performance may be pervasive, common, and even normative, but it may still be inadequate for the needs of society and below the level of skills that would be judged as desired and appropriate by parents and educators.

Table 1
Benchmark Goals and Timelines for a Trajectory of Progress Toward High-Stakes Reading Outcomes

Timeline	Measure	Benchmark Goal for a Trajectory of Progress	May Need Intensive Instructional Support
Winter, Kindergarten	Onset Recognition Fluency	25 – 35 Onsets Correct per Minute	Below 10 Onsets Correct per Minute
Spring, Kindergarten	Phoneme Segmentation Fluency	35 – 45 Phonemes Correct per Minute	Below 10 Phonemes Correct per Minute
Winter, First Grade	Nonsense-Word Fluency	50 Letter-Sounds Correct per Minute	Below 30 Letter-Sounds Correct per Minute
Spring, First Grade	CBM Oral Reading Fluency	40 Words Correct per Minute in grade-level material	Below 10 Words Correct per Minute in grade-level material
Spring, Second Grade	CBM Oral Reading Fluency	90 Words Correct per Minute in grade-level material	Below 50 Words Correct per Minute in grade-level material
Spring, Third Grade	CBM Oral Reading Fluency	110 Words Correct per Minute in grade-level material	Below 70 Words Correct per Minute in grade-level material

A second problem with using an entirely normative basis to establish benchmark goals is, suppose that intervention, instruction, and curricular improvements actually work. After all, the intent of a goal is to provide a target for all children to attain. But, if we have a normative-based target, and we are effective in reaching the target, the target will necessarily move. No matter how effective our instruction, 50% of children will still be below the middle performance. Although normative comparisons can help to interpret and understand goals, they provide a problematic basis upon which to establish a goal.

The present study investigated the decision-making utility of a prevention-oriented assessment and intervention system that uses fluency-based indicators of foundational skills of early reading. Specifically, we examined the following research questions:

1. What is the decision-making utility of the DIBELS benchmark goals in the context of a district engaged in a school-wide educational reform effort targeting phonological awareness and alphabetic principle skills?
2. What is the decision-making utility of the first grade CBM ORF benchmark goal with respect to continued progress toward reading outcomes judged desirable and appropriate?
3. What is the strength of the relation between CBM ORF and high-stakes reading outcomes?
4. What level of proficiency on CBM ORF predicts successful attainment of the state standard? What level of performance predicts failure?

Method

Setting and Participants

Participants were four cohorts of students from kindergarten through Grade 3 from six elementary schools in a fast-growing (i.e., approximately 5% population growth per year), urban district of the Pacific Northwest. The kindergarten 1998-99 / first grade 1999-2000 cohort provided information on the linkage from DIBELS PSF in spring of kindergarten to DIBELS NWF in winter of first grade ($n = 302$) to CBM ORF in spring of first grade ($n = 378$). (See Table 2 for a listing of cohort size by analysis.) The total district K-12 enrollment was 5,246 students; five of the six elementary schools qualified for Title I services with the percentage of children receiving free and reduced lunch ranging from a low of 37% to a high of 63% in the respective schools. Within the district, 10% of students were considered minority; 18% of total children enrolled were considered at or below the poverty range. All six schools in the district were participating in a model demonstration project, Accelerating Children's Competence in Early Literacy-Schoolwide (ACCEL-S), funded by the U.S. Department of Education and designed to improve the reading of all students in Grades K-3 (Simmons, Kame'enui, & Good, 1998).

Table 2
Subject Cohorts and Variables for Literacy Linkages

Linkage	Cohort	Variables	<u>n</u>	Mean	<u>SD</u>
1	Kindergarten 1999-2000	OnRF Winter K	353	27.02	14.16
		PSF Spring K	353	45.72	16.09
2	Kindergarten 1998-99 & First Grade 1999-2000	PSF Spring K	302	41.26	18.90
		NWF Winter 1	302	52.56	28.51
3	Kindergarten 1998-99 & First Grade 1999-2000	NWF Winter 1	378	51.66	28.56
		ORF Spring 1	378	55.67	37.69
4	First Grade 1998-99 & Second Grade 1999-2000	ORF Spring 1	342	34.23	29.61
		ORF Spring 2	342	91.22	38.29
5	Third Grade 1999-2000	ORF Spring 3	364	113.25	37.04
		OSA Spring 3	364	212.83	13.69

Note. OnRF = Onset Recognition Fluency, PSF = Phoneme Segmentation Fluency, NWF = Nonsense Word Fluency, ORF = Oral Reading Fluency, OSA = Oregon Statewide Assessment.

Measures

To evaluate the role of fluency in the development of foundation skills in beginning reading, we utilized three types of measures: (a) fluency-based measures of early literacy (i.e., Dynamic Indicators of Basic Early Literacy Skills (DIBELS) (Kaminski & Good, 1996; DIBELS measures, procedures, and support are available at idea.uoregon.edu/~dibels); (b) a curriculum-based measure of oral reading fluency (i.e., Test of Reading Fluency) (Children's Educational Services, 1988); and (c) a high-stakes measure of comprehensive reading achievement (Oregon Statewide Assessment). Each measure is described below.

Fluency-Based Measures: Dynamic Indicators of Basic Early Literacy Skills

DIBELS Onset Recognition Fluency (OnRF). DIBELS OnRF is a standardized, individually administered measure of phonological awareness that assesses a child's ability to recognize and produce the initial sound in an orally presented word (Kaminski & Good, 1996, 1998; Laimon, 1994). The examiner presents four pictures to the child, names each picture, and then asks the child to identify (i.e., point to or say) the pictures that begins with the sound produced orally by the examiner. The child is also asked to orally produce the beginning sound for an orally presented word that matches one of the given pictures. The examiner calculates the amount of time taken to identify/produce the correct sound and converts the score into the number of onsets correct in a minute. Alternate form reliability of the OnRF measure is .72 in January of kindergarten (Good et al., in preparation). While that level of reliability is low with

respect to standards for educational decision-making (e.g., Salvia & Ysseldyke, 2001), it is remarkable in a one-minute measure – especially one that can be repeated. By repeating the assessment 4 times, the resulting average would have a reliability of .91 (Nunnally, 1978). The concurrent criterion related validity of OnRF with DIBELS PSF is .48 in January of kindergarten, and .36 with the Woodcock-Johnson Psycho-Educational Battery readiness cluster score (Good et al., in preparation). The predictive validity of OnRF with respect to spring first-grade reading on CBM ORF was .45, and .36 with the Woodcock-Johnson Psycho-Educational Battery total reading cluster score 65 (Good et al., in preparation).

DIBELS Phonemic Segmentation Fluency (PSF). The PSF measure is a standardized, individually administered, test of phonological awareness (Kaminski & Good, 1996). The PSF measure assesses a student's ability to segment three and four phoneme words into the individual phonemes fluently. The PSF measure has been found to be a good predictor of later reading achievement and is intended for use with students from the winter of kindergarten to the middle of first grade (Kaminski & Good, 1996). The PSF task is administered by the examiner orally presenting words of three to four phonemes. It requires the student to produce verbally the individual phonemes for each word. For example, the examiner would say “sat” and the student would need to say “/s/ /a/ /t/” to receive three possible points for the word. After the student responds, the examiner presents the next word and the number of correct phonemes produced within one minute determines the final score. The two-week, alternate-form reliability for the PSF measure was .88 (Kaminski & Good, 1996) and the one-month, alternate-form reliability was .79 in May of kindergarten (Good et al., in preparation). Concurrent criterion validity of PSF is .54 with the Woodcock-Johnson Psycho-Educational Battery readiness cluster score in spring of kindergarten (Good et al., in preparation). The predictive validity of spring, kindergarten PSF with (a) winter first grade DIBELS NWF was .62, (b) spring first grade spring Woodcock-Johnson Psycho-Educational Battery total reading cluster score was .68, and (c) spring of first grade CBM ORF was .62 (Good et al., in preparation).

DIBELS Nonsense Word Fluency (NWF). The DIBELS NWF measure is a standardized, individually administered, test of letter-sound correspondence and of the ability to blend letters into words in which letters represent their most common sounds (Kaminski & Good, 1996). The student is presented with a 8.5” x 11” sheet of paper with randomly ordered VC and CVC nonsense words (e.g., sig, rav, ov) and asked to produce verbally the individual letter sound of each letter or verbally produce, or read, the whole nonsense word. For example, if the stimulus word is “vaj” the student could say /v/ /a/ /j/ or say the word /vaj/ to obtain a total of three letter sounds correct. The student is allowed one-minute to produce as many letter-sounds as he/she can, and the final score is the number of letter-sounds produced correctly in one minute. Because the measure is fluency based, students receive a higher score if they are phonologically recoding the word and receive a lower score if they are providing letter sounds in isolation. The one-month, alternate-form reliability for NWF in January of first grade was .83 (Good et al., in preparation). The concurrent criterion-validity of DIBELS NWF with the Woodcock-Johnson Psycho-Educational Battery-Revised readiness cluster score was .36 in January and .59 in February (Good et al., in preparation). The predictive validity of DIBELS NWF in January of first grade with (a) CBM ORF in May of first grade was .82, (b) CBM ORF in May of second grade was .60, (c) Woodcock-Johnson Psycho-Educational Battery total reading cluster score was .66 (Good et al., in preparation).

Curriculum-Based Measurement of Oral Reading Fluency (CBM ORF). Three passages from the Grade 3 screening and Level C progress monitoring passages of the Test of Reading Fluency (TORF) (Children's Educational Services, 1987) were used to assess oral reading fluency. The TORF is a standardized set of passages and administration procedures designed to (a) identify children who may need further intensive assessment, and (b) measure growth in reading skills (Children's Educational Services, 1987, p. 1). Passages were calibrated for each grade level, and student performance is measured by having students read each of three passages aloud for one minute. Words omitted, substituted, and hesitations of more than three seconds are scored as errors. Words self-corrected within three seconds are scored as accurate. The median correct words per minute from the three passages were selected as the oral reading fluency rate.

A series of studies have confirmed the technical adequacy of the TORF. Test-retest reliabilities of elementary-aged students ranged from .92 to .97; alternate-form reliability of different reading passages drawn from the same level ranged from .89 to .94 (Tindal, Marston, & Deno, 1983). Criterion-related validity studied in eight separate studies in the 1980s reported coefficients ranging from .52-.91 (Good & Jefferson, 1998).

Standardized Measure of Comprehensive Reading Achievement

Oregon Statewide Assessment - Reading/Literature (OSA). The OSA in reading/literature is a standardized achievement test developed by panels of teachers in concert with a research and development company (Oregon Department of Education, 2000). The test uses a multiple-choice format, and the primary purpose of the test is to assess the achievement level of individual students and compare the achievement with performance standards established by the Oregon State Board of Education at each Benchmark level (i.e., Grades 3, 5, 8 and 10). The OSA uses a multiple-form design (i.e., Forms A-D); the internal consistency reliability (KR – 20) calculated across four alternate forms for Grade 3 Reading/Literature was .95 (Oregon Department of Education, 1996). The third-grade mean was 206 with a standard deviation of 12.16 in total reading. For a school with 100-120 students, the mean standard error was .31.

All students in Grade 3 are tested routinely by the school district in the spring and given approximately 90 minutes to complete the 56 items on the reading test. However, students are given more time if needed. Students read six passages that range in length, variety, and cover a broad range of topics. The scale for the multiple-choice test is considered a “growth scale” and each point on the scale is an equal distance from the previous point on the scale so changes can be charted and viewed as comparable from year to year. The scale ranges from 150 to 300. A score of 201 or above is described as “meets expectations” for the Grade 3 benchmark and a score of 215 or above is described as “exceeds expectations” for the Grade 3 benchmark. A score below 201 is described as “does not meet expectations.” In the 2000 sample of 38,730 students, 18% did not meet the Grade 3 benchmark, 30% met the benchmark, and 52% exceeded the benchmark. Results of the assessment are published and disseminated on a school-by-school basis.

Results

A series of linked, short-term, longitudinal studies of four cohorts was used to examine the strength of relations and performance probabilities among foundational reading measures and

a third-grade high-stakes reading assessment. Performance linkages were examined in the 1998-99 and 1999-2000 academic years. The number of children in each cohort, their grade and academic year placement, and descriptive statistics are reported in Table 2. To the greatest extent possible, all students in the district were included. When the linkage extended across academic years, the number of students with complete information is reported. Spring second-grade performance for the third grade 1999-2000 cohort was not available, so the second- to third-grade linkage illustrated in Figure 1 was not examined in this study.

The strength of the linkages between subsequent skills are frequently and traditionally examined with correlation coefficients and percent of variance explained. The correlation between subsequent skills, and the percent of variance explained in subsequent skills are summarized in Table 3. As indicated, for this district, the correlations between earlier and later skills ranged from .34 to .82. The variance explained ranged from 12 percent to 67 percent. In addition to the correlation and percent of variance explained, the purpose of this paper was to examine the utility of the benchmark goals established for DIBELS and CBM ORF measures.

Table 3
Strength of Literacy Linkages from Kindergarten through Third-Grade High-Stakes Outcomes

Earlier Benchmark Goal / Next Benchmark Goal	<u>n</u>	<u>r</u> *	Percent of variance explained	Percent of students who need intensive instructional support who attained next benchmark goal	Percent of students who reach the earlier benchmark goal and who attained the next benchmark goal
OnRF Winter K / PSF Spring K	353	.34	12	29%	91%
PSF Spring K / NWF Winter 1	302	.38	14	11%	55%
NWF Winter 1 / ORF Spring 1	378	.78	60	9%	90%
ORF Spring 1 / ORF Spring 2	342	.82	67	0%	97%
ORF Spring 3 / OSA Spring 3	364	.67	45	28%	96%

Note. OnRF = Onset Recognition Fluency, PSF = Phoneme Segmentation Fluency, NWF = Nonsense Word Fluency, ORF = Oral Reading Fluency, OSA = Oregon Statewide Assessment.

*All relations significant, $p < .001$.

Utility of DIBELS Onset Recognition Fluency Goal

The intent of a benchmark goal is to specify a level of performance where the odds of attaining subsequent goals are in the teachers' (and children's) favor. The relation between DIBELS Onset Recognition Fluency (OnRF) in winter of kindergarten and DIBELS Phoneme Segmentation Fluency (PSF) in spring of kindergarten is illustrated in Figure 3. Students portrayed in this figure were in kindergarten during the 1999 – 2000 academic year.

The vertical line at OnRF of 25 represents the winter-of-kindergarten benchmark goal. Of the 188 kindergarten students attaining the winter of kindergarten OnRF benchmark goal, 172 (91%) attained the PSF benchmark goal in spring of kindergarten. However, of the 24 students who scored below 10 on OnRF, only 7 (29%) attained the spring goal. Obtaining an OnRF score

between 10 and 25 in winter of kindergarten resulted in a less clear prediction. For teachers in the beginning months of kindergarten, a goal of 25 on OnRF by winter represents a level of phonological awareness where the odds are in their favor of reaching the spring kindergarten goal. Thus, OnRF has decision-making utility as an instructional goal, and as a basis for evaluating student progress toward reading outcomes.

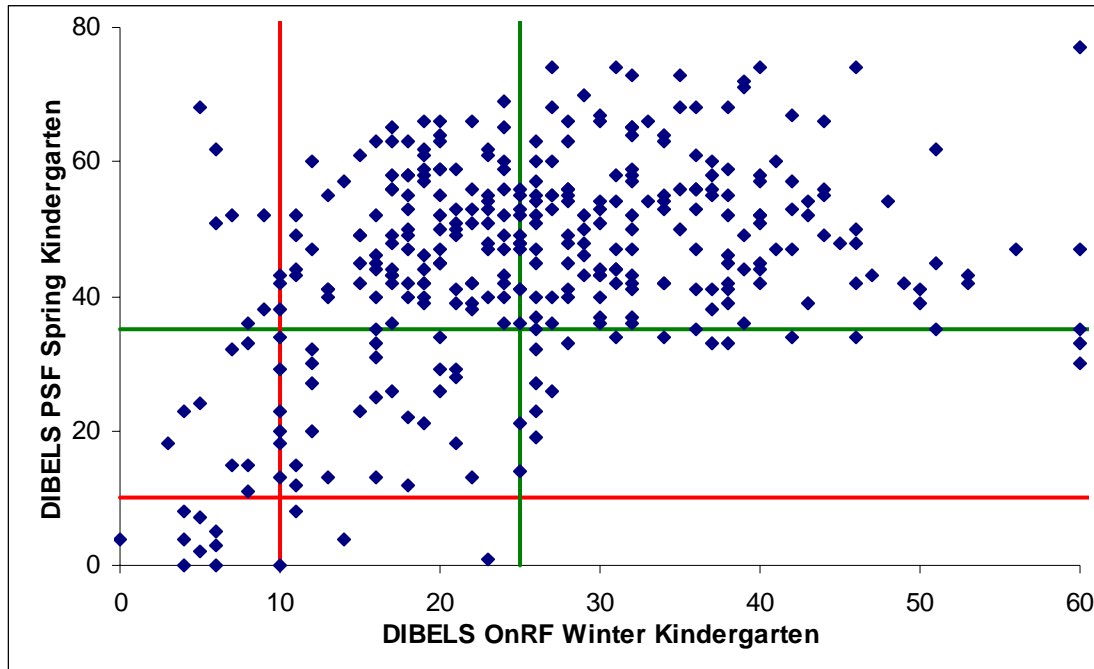


Figure 3. Linkage between DIBELS Onset Recognition Fluency in winter of kindergarten and DIBELS Phoneme Segmentation Fluency in spring of kindergarten.

Utility of DIBELS Phoneme Segmentation Fluency Goal

The linkage between May of kindergarten DIBELS PSF and winter of first-grade DIBELS NWF is illustrated in Figure 4. Students in this figure were enrolled in kindergarten in the 1998-99 academic year, and were in first grade for the 1999-2000 academic year. The vertical line at 35 in the spring of kindergarten is the benchmark goal for DIBELS PSF. In the spring of kindergarten, 201 students met the goal and 110 (55%) of those students later attained the winter first-grade benchmark goal on DIBELS NWF. Of the 19 students who scored below 10 on DIBELS PSF in spring of kindergarten, only 2 (11%) later attained the winter first-grade benchmark goal. The vertical line at 10 on DIBELS PSF indicates a level where intensive instructional support will probably be necessary to attain later reading goals. The prediction of reading outcomes is not clear for students scoring between 10 and 35 on PSF in spring of kindergarten.

Utility of DIBELS Nonsense Word Fluency Goal

Figure 5 illustrates the linkage between winter of first-grade DIBELS NWF and the CBM ORF in the spring of first grade. These students were assessed during the 1999-2000 academic year. The vertical line at 50 in Figure 4 corresponds to the benchmark goal for

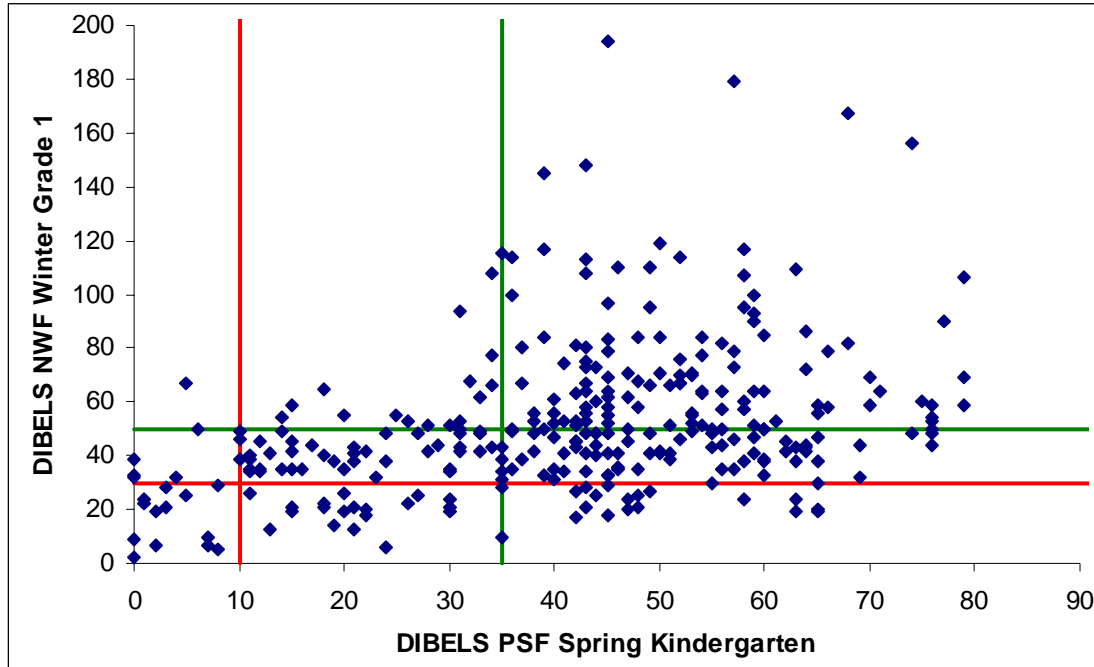


Figure 4. Linkage between spring kindergarten phonological awareness on DIBELS Phoneme Segmentation Fluency and winter first-grade alphabetic principle on DIBELS Nonsense Word Fluency.

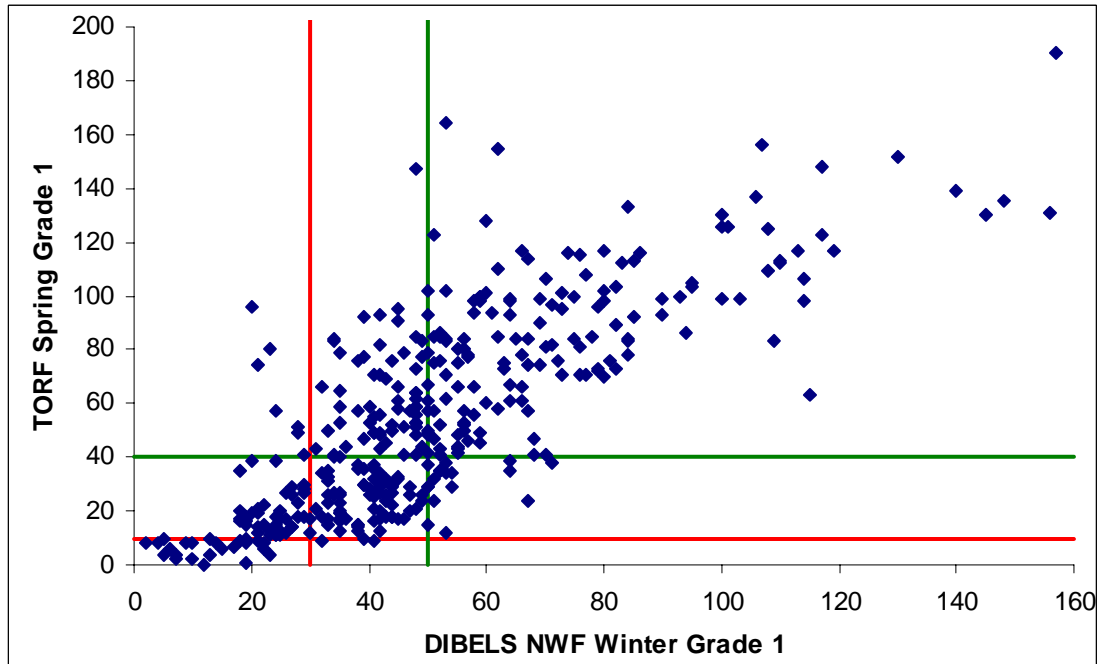


Figure 5. Linkage between winter first-grade DIBELS Nonsense Word Fluency and spring-of-first-grade reading on the Test of Reading Fluency.

winter of first grade. In this sample, 169 students reached the winter goal and 152 of those students (90%) subsequently attained the spring of first-grade reading benchmark goal. Of the 74 students scoring below 30 on DIBELS NWF in winter of first grade, only 7 (9%) attained the spring of first-grade reading goal. Thus, the vertical line at 30 indicates a level where intensive instructional support will probably be needed for a student to attain the first-grade reading goal.

Utility of CBM Oral Reading Fluency Grade 1 Goal

The linkage between first-grade CBM ORF for students in first grade in the 1998-99 academic year and second-grade CBM ORF outcomes in the 1999-2000 academic year is illustrated in Figure 6. The vertical line at 40 corresponds to the first grade benchmark goal. Of the 98 students who reached the first-grade benchmark, 95 or 97% attained the second-grade benchmark goal. Thus, the first grade benchmark goal of 40 on CBM ORF appears to have utility as a goal that predicts continued reading progress. Of the 51 students reading below 10 words in spring of first grade, none attained the second-grade benchmark goal. Thus, a score below 10 on CBM ORF in spring of first grade appears to have utility as a level where intensive instructional support will probably be needed if the student is going to attain the second-grade goal. Students scoring between 10 and 39 on CBM ORF in spring of first grade were less clearly predictable. They may need additional instructional support to attain second-grade outcomes.

Utility of CBM ORF Grade 3 Goal

The linkage between May of third-grade CBM ORF and third-grade performance on the OSA is illustrated in Figure 7. Students in this figure were enrolled in third grade in the 1999-2000 academic year. The two horizontal lines correspond to the state of Oregon standards of "meets expectations" at a score of 201, and "exceeds expectations" at a score of 215 on the OSA. A score below 201 on the OSA corresponds to "does not meet expectations."

The vertical line at 110 corresponds to a CBM ORF benchmark goal for May of third grade where students are likely to meet or exceed expectations on the OSA. Of the 198 students who attained the May of third-grade goal, 191 or 96% were rated as "meets expectations" or "exceeds expectations" on the OSA. For students reading between 70 and 110 on the CBM ORF passages, the likelihood of meeting expectations on the OSA was less clear. Students scoring

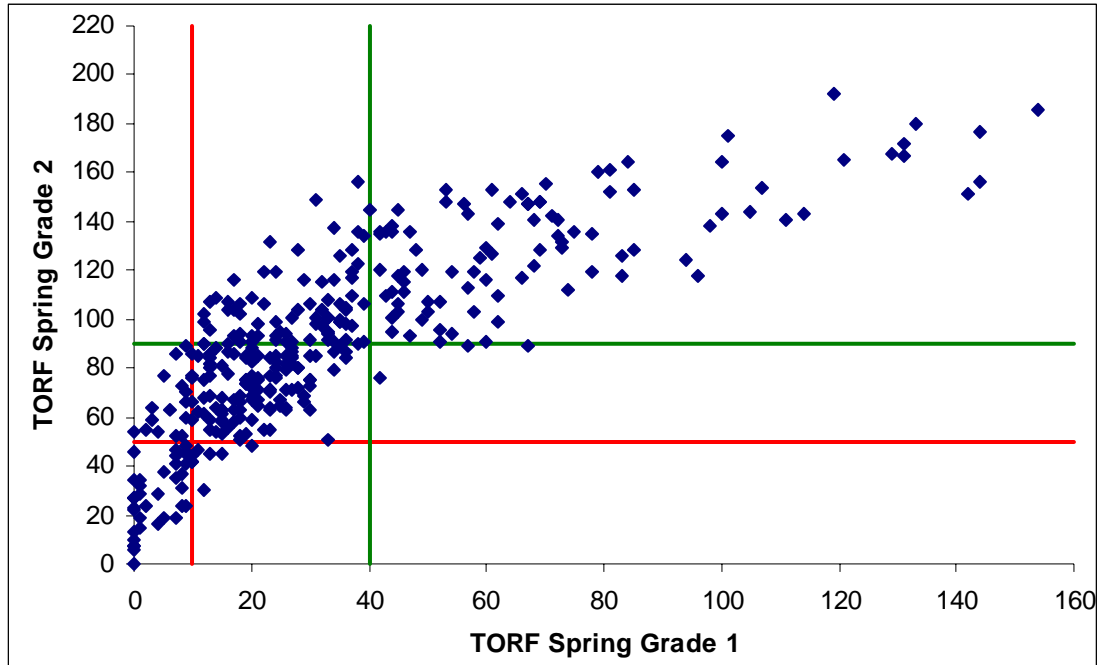


Figure 6. Linkage between first- and second-grade reading on the Test of Reading Fluency passages in the Spring.

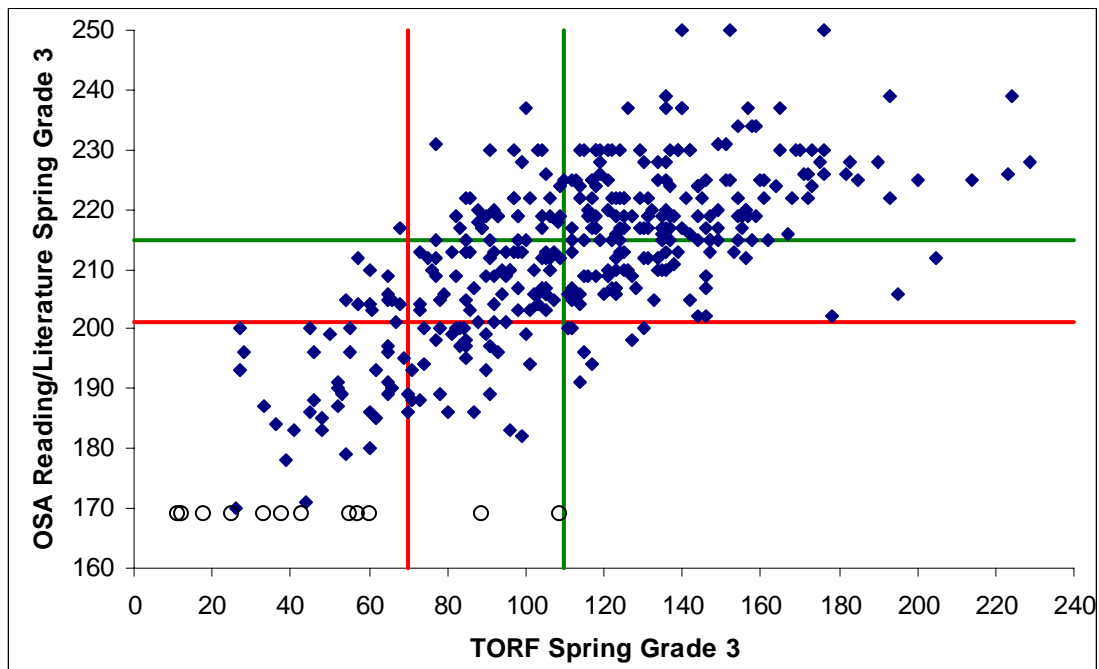


Figure 7. Linkage between Curriculum-Based Measurement spring reading in third grade and passing the Oregon Statewide Assessment.

below 70 were unlikely to meet expectations on the OSA. Of the 46 students who scored below 70, only 13 or 28% were rated as “meets expectations” on the OSA. Thus, the vertical line at 70 correct words per minute corresponds to the need for intensive instructional support. A circle on the figure indicates one of the 12 students for whom a standard OSA score was not available. These students either (a) were administered a modified OSA and rated as “does not meet

expectations” ($n = 4$), or (b) were not administered the OSA ($n = 8$). Our hypothesis is that standard OSA scores were not randomly missing, but that a missing score reflected a prediction, formal or informal, by school personnel that the student would not pass the OSA. If a standard OSA score was not available, a circle was plotted at an OSA score of 169, consistent with a prediction that the student would not meet expectations. Students missing a standard OSA score were not included in calculating correlation coefficients or utility percentages.

Discussion and Instructional Implications

Over the past decade, schools have experienced both the rhetoric and reality of high-stakes assessment. The instruments and technology of assessment are being summoned with increased frequency and intensity to assess all students’ level of achievement with respect to high-stakes reading outcomes. The existing measures and assessment methodologies are ill-prepared to meet one of the most critical purposes of assessment--to forecast attainment of high-stakes outcomes early enough to inform instruction and alter learning trajectories. In this article, we introduced a conceptual and procedural measurement model using fluency-based indicators of foundational reading skills and examined its utility for predicting future performance and informing instruction.

Utility of DIBELS Benchmark Goals

One purpose of this study was to examine the decision-making utility of the DIBELS benchmark goals in the context of a district engaged in a school-wide educational reform effort targeting phonological awareness and alphabetic principle skills. With one possible exception, this study provides strong support for the utility of the benchmark goals. Students who attained the earlier benchmark goal were highly likely ($> 90\%$) to attain the subsequent literacy benchmark goal. The exception to this pattern of findings was that support for the utility of the spring-of-kindergarten PSF benchmark goal was less strong (55% attained the subsequent goal).

One hypothesis for the lower utility of the DIBELS PSF measure in this study is that the measure has lower predictive validity and less utility as a benchmark goal. However, the results of this study were inconsistent with prior research on the predictive validity of DIBELS PSF (Good et al., in preparation; Johnson, 1996). A second plausible hypothesis for the lower utility of DIBELS PSF and the difference in utility of the DIBELS PSF measure compared to prior research is that the differences in utility are due to the differences in the instructional context. In this district, students received the benefit of a schoolwide educational reform effort targeting phonological awareness and alphabetic principle skills. In the 1998-99 academic year, the reform effort primarily targeted kindergarten instruction and support. All kindergarten teachers in the district received in-service training in research-based practices in early literacy. They adopted curricula that were research based, and supplemented their curricula with interventions targeting the big ideas of early literacy as needed. The district invested additional instructional and curricular resources to ensure that all kindergarten children learn phonological awareness and alphabetic principle skills. The finding that 69% of all kindergarten children reached the spring kindergarten PSF benchmark goal supports the strength of the kindergarten curriculum and instruction. In comparison, only 21% of all kindergarten children had reached the spring-of-kindergarten benchmark goal in another research site not engaged in schoolwide reform efforts targeting phonological awareness (Good et al., in preparation).

According to this instructional context hypothesis, the instructional effort and support provided in kindergarten on phonological awareness were effective in supporting many children to the spring-of-kindergarten PSF benchmark goal, but the instruction on alphabetic principle skills provided in kindergarten and first grade was not sufficient to support many of those children to attain the winter-of-first grade NWF benchmark goal. In this discussion, it is important to keep in mind that the utility of a benchmark goal is not based just on predictive validity. The prevention-oriented assessment and intervention system described here builds on research-based “big ideas” of reading acquisition: (a) phonological awareness, (b) alphabetic principle, (c) and accuracy and fluency with connected text (see Figure 1). In this study, the lower utility of DIBELS PSF resulted from students with a pattern of timely attainment of phonological awareness skills, but insufficient alphabetic principle skills in time to change first-grade reading outcomes. The implications of this pattern for instructional effort and reform are direct.

Instructional Implications of Zones of Performance

The performance linkages in the measurement model based on the research-based big ideas in early literacy provide four performance zones relevant to system-wide instructional decisions (see Figure 8). The zones of performance for Figure 3, the linkage between spring kindergarten PSF and winter first grade NWF, illustrate these instructional implications. Similar interpretation would be appropriate for each of the linked steps in the prevention-oriented assessment and intervention system described here. In Figure 8, Zone A represents students who achieved benchmark goals on an earlier skill at an earlier time and who then achieved the benchmark goal on a later skill at a later time. For each of the linkages examined, students in Zone A are progressing on a trajectory that results in successful reading outcomes. Students who follow this pattern for each of the benchmark goals in the model of reading acquisition would be on track for successful performance on high-stakes reading outcome measures. Thus, Zone A represents the desired pattern of performance and the goal of effective instruction.

The remaining three performance zones illustrated in Figure 8 provide information about students whose performance trajectories indicate weak “links” or instructional areas that may jeopardize successful reading outcomes. In some ways, instructing students toward reading outcomes is similar to running a relay race. There are critical legs that contribute to the overall outcomes. If students pass from one leg to the next behind in foundational skills, the high-stakes outcome is jeopardized. A weak leg of the academic race can potentially be recovered with a strong compensatory effort later in the race; however, prior research documents that the odds of this occurring decrease with time (e.g., Juel, 1988). Students who achieved the earlier benchmark goal but who did not achieve the later benchmark goal would be plotted in Zone B. This pattern tells us the instructional advantage established earlier was not sustained. Students who did not achieve the earlier benchmark goal but for whom a strong instructional effort was effective in achieving the subsequent benchmark goal are plotted in Zone C. Finally, students plotted in Zone D did not achieve either the earlier or later benchmark goal. The reading progress of students in Zone D is not sufficient to make a confident prediction of reading outcomes. To the extent that students are in the lower left quadrant of Zone D, the likelihood of attaining reading outcomes decreases. By using the system of linkages from kindergarten through third grade, a school can identify strengths and weaknesses in their instructional support.

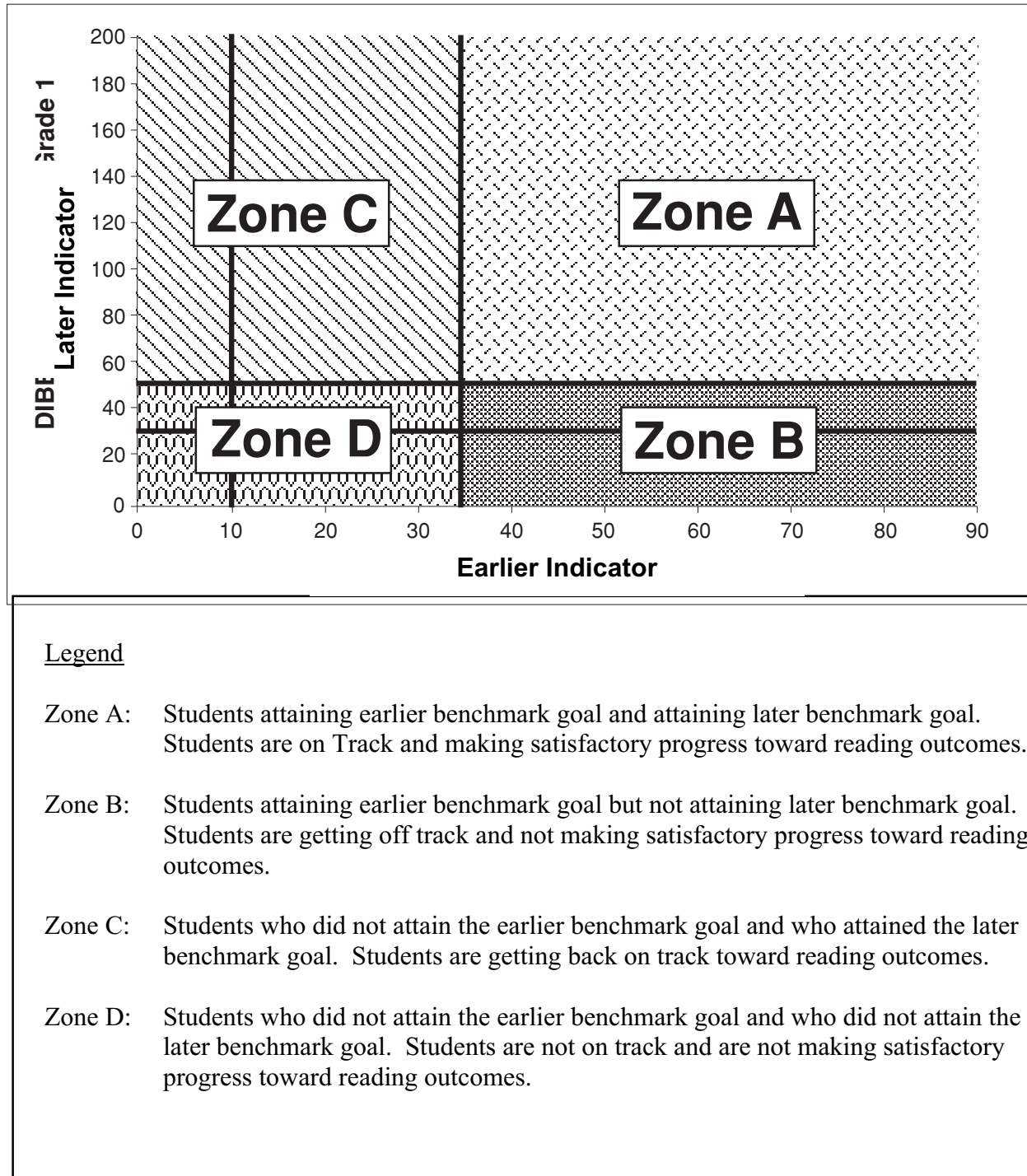


Figure 8. Instructionally interpretable zones of performance in a fluency-based model of the acquisition of early literacy skills and reading proficiency.

When instruction and assessment are tightly linked, predictive validity alone may not provide a sufficient basis to evaluate the utility of the measures. For example, if a school district focused their instruction and curriculum on attaining the benchmark goals for all students and most or all students were plotted in Zone A, then the correlation between earlier and later performance (i.e., predictive validity) would be essentially zero. Similarly, when the instructional

context is such that many students are plotted in Zone B or Zone C, lower predictive validity correlations will be found, but the measures may have utility for identifying strengths and weakness in the curriculum or instruction. In sum, a measurement system has utility to the extent the measures inform instruction and contribute to reading outcomes.

Utility of CBM ORF Benchmark Goals

A second purpose of this study was to examine the utility of the first grade CBM ORF benchmark goal with respect to continued progress toward reading outcomes. The first-grade outcomes were strongly predictive of continued progress in second grade and consistent with desired second-grade outcomes. Of particular concern are the students plotted in Zone D who did not achieve the first-grade reading benchmark and did not attain the second-grade reading benchmark goal. In general, the reading progress of students in Zone D is not sufficient to make a confident prediction of reading outcomes. To the extent that students are in the lower left quadrant of Zone D, the likelihood of attaining reading outcomes decreases. For these students, the single best way to increase second-grade reading outcomes is to attain the spring-of-first-grade benchmark reading goal on CBM ORF.

A third purpose of this study was to examine the strength of the relation between fluency with connected text as measured by CBM ORF and high-stakes reading outcomes. The results of this study support accuracy and fluency with connected text as an important foundation for reading competence. Students who read grade-level material at a rate of 110 words correct per minute or better were likely to meet or exceed expectations on the Oregon Statewide Assessment. Students who were able to read less than 70 words correct per minute on grade-level material were not likely to meet expectations on the Oregon Statewide Assessment.

Implications for Further Research

As we continue to explore and refine measurement methods to inform instruction and pre-empt reading failure on high-stakes outcomes, we recognize the need for systematic investigation in the following areas. First, as with most studies, longer-term followup with students as they progress into higher grades is clearly important to assess the utility of early measures to forecast long-term outcomes. Specifically, would performance in Grade 1 predict Grade 5 performance on high-stakes outcomes and beyond? In addition to studying the ability of early measures to forecast long-term performance, further research is necessary to study the generalizability of findings of linked longitudinal studies to true longitudinal studies. We are in the process of assessing the performance of three separate cohorts of students longitudinally to examine the linkages of the model in longitudinal performance across cohorts (Simmons & Kame'enui, 1998).

A crucial area of need for additional research is an examination of the district-to-district variability in the patterns of linkages between early literacy and reading skills, and an examination of the important features of the instructional context that affect the patterns. The instructional context of the school-wide educational reform effort is consistent with the obtained pattern of linkages, and informative about the need for further instructional modifications. Further research is needed to examine the range of patterns for various districts.

Fluency was a common denominator of the measures used to assess foundational reading skills. As Fuchs et al., (this issue) reported, fluency-based measures of connected text were better

discriminators than accuracy-based measures of connected text and correlated more strongly with measures of general reading competence. Analyses comparing fluency to accuracy measures for early reading indicators are emerging, yet incomplete. Our preliminary analysis of phonemic segmentation proficiency of kindergarten students indicated a strong correlation between DIBELS PSF and the Yopp-Singer (Yopp, 1995) measure of phonemic segmentation ($r = .77$) (Kame'enui, Simmons, & Good, 2000). Correlations of such strong magnitude support the use of one-minute, fluency-based measures that efficiently and reliably document phonemic awareness skill and progress. Nevertheless, further research is necessary to replicate this finding and extend research into other areas of early reading.

It is important to be mindful that the reading performance documented in this study took place in an innovative environment with a strong focus on research-based practices and reading improvement. From central administration, to school administrators, to classroom teachers, and educational assistants, the focus of the district was to ensure that each child would read by Grade 3. The utility of the DIBELS and CBM ORF benchmark goals as an instructional target that would change outcomes has contributed to the educational reform effort by focusing instructional resources, targeting areas of instructional strength and need, and tracking individual student progress toward key benchmark goals. The strong linkages of performance for students who met benchmark on early indicators and likewise achieved later benchmark goals has contributed to a change in reading outcomes for the district.

At the other end of the prediction continuum is the ability of early reading indicators to portend subsequent reading difficulty. Our findings consistently indicated that students who scored low on one indicator were at serious risk of not attaining acceptable levels of performance on subsequent measures. For these students, the goal must be to ruin the prediction; that is, to alter proactively the instruction and learning conditions sufficient so that where children began does not forecast where they will end. It is for this reason that our focus must be on a prevention-oriented assessment and intervention system with utility for making instructional decisions that change student outcomes.

Results of this study underscore the utility of fluency-based indicators of foundational reading skills to inform instructional decisions early enough to change outcomes before reading problems become too large and established. With strong and remarkable consistency, the performance linkages across measures supported the utility of early measures to predict later performance and the hypothesized importance and relation of fluency of foundational skills to later reading outcomes (Logan, 1997a; 1997b). In an era of high-stakes assessment, an assessment system that can be used in concert with instruction to prevent pervasive and enduring long-term reading difficulty holds extraordinary potential. Future studies must replicate and extend the current findings in more diverse settings, over longer periods of time, and with a broader array of high-stakes outcomes. The opportunity to apply, extend, replicate, and refine what has been learned in this study is of significant relevance and promise as we continue to determine the elements of an assessment and intervention system necessary to improve reading outcomes for each and all.

Author Note

We express our sincere appreciation to the administrators, teachers, educational assistants, and students of the Bethel School District for their commitment to reading achievement and for allowing us to study Schoolwide reading improvement and to learn from the process.

The contents of this document were developed with the support of the Office of Special Education Programs, U.S. Department of Education, under Contract Number H324M980127. This material does not necessarily represent the policy of the U.S. Department of Education, and the material is not necessarily endorsed by the federal government.

References

- Ackerman, P. L. (1987). Individual differences in skill reading: An integration of psychometric and information processing perspectives. Psychological Bulletin, 102, 3-27.
- Adams, M. J. (1990). Beginning to read: Thinking and learning about print. Cambridge, MA: MIT Press.
- Bond, L., Roeber, E., & Connealy, S. (1998). Trends in state student assessment programs. Washington, D.C.: Council of Chief State School Officers.
- Bus, A. G., & van IJzendoorn, M. H. (1999). Phonological awareness and early reading: A meta-analysis of experimental training studies. Journal of Educational Psychology, 91, 403-414.
- Carnine, D. (1997). Instructional design in mathematics for students with learning disabilities. Journal of Learning Disabilities, 30, 130-131.
- Carnine, D. W. (2000). A consortium for Evidence in Education (CEE). Unpublished manuscript.
- Chall, J. S. (1983). Stages of reading development. New York: McGraw-Hill.
- Children's Educational Services. (1987). Test of Reading Fluency. Minneapolis, MN: Author.
- Deno, S. L., Mirkin, P., & Chiang, B. (1982). Identifying valid measures of reading. Exceptional Children, 49, 36-45.
- Drucker, P. F. (1993). The rise of the knowledge society. The Wilson Quarterly, 17, 52-72.
- Ehri, L. C., & McCormick, S. (1998). Phases of word learning: Implications for instruction with delayed and disabled readers. Reading and Writing Quarterly, 14, 135-163.
- Elmore, R. F. (1996). Getting to scale with good educational practice. Harvard Educational Review, 66(1), 1-26.
- Francis, D. J., Shaywitz, S. E., Stuebing, K. K., Shaywitz, B. A., & Fletcher, J. M. (1994). Measurement of change: Assessing behavior over time and within a developmental context. In G. R. Lyon (Ed.), Frames of reference for the assessment of learning disabilities: New views on measurement issues. Baltimore, MD: Paul H. Brookes.
- Fuchs, L. S., & Deno, S. L. (1991). Paradigmatic distinctions between instructionally relevant measurement models. Exceptional Children, 58, 232-243.
- Fuchs, L. S., Fuchs, D., Hamlett, C. L., Walz, L., & Germann, G. (1993). Formative evaluation of academic progress: How much growth can we expect? School Psychology Review, 22, 27-48.
- Fuchs, L. S., & Fuchs, D. (1999). Monitoring student progress toward the development of reading competence: A review of three forms of classroom-based assessment. School Psychology Review, 28, 659-671.
- Fuchs, L. S., Fuchs, D., Hosp, M. K., & Jenkins, J. (this issue). Oral reading fluency as an indicator of reading competence: A theoretical, empirical, and historical analysis. Scientific Studies of Reading.

Good, R. H. I., & Jefferson, G. (1998). Contemporary perspectives on Curriculum-Based Measurement Validity. In M. R. Shinn (Ed.), Advanced applications of Curriculum-Based Measurement (pp. 61-88). New York: Guilford.

Good, R. H., Kaminski, R. A., Shinn, M. R., Bratten, J., Shinn, M. M., & Laimon, D. (in preparation). Technical adequacy of Dynamic Indicators of Basic Early Literacy Skills. Research Report #7. Early Childhood Research Institute on Measuring Growth and Development, University of Oregon.

Good, R. H., Simmons, D. C., & Smith, S. (1998). Effective academic interventions in the United States: Evaluating and enhancing the acquisition of early reading skills. School Psychology Review, 27, 45-56.

Green, P. C., & Sireci, S. G. (1999). Legal and psychometric policy considerations in the testing of students with disabilities. Journal of Special Education Leadership, 12, 21-29.

Hasbrouck, J. E., & Tindal, G. (1992, Spring). Curriculum-based oral reading fluency norms for students in grades 2 through 5. Teaching Exceptional Children, pp. 41-44.

Johnson, D. S. (1996). Assessment for the prevention of early reading problems: Utility of dynamic indicators of basic early literacy skills for predicting future reading performance. Unpublished Doctoral Dissertation, University of Oregon, Eugene, OR.

Juel, C. (1988). Learning to read and write: A longitudinal study of 54 children from first through fourth grades. Journal of Educational Psychology, 80, 437-447.

Kame'enui, E. J., & Carnine, D. W. (Eds.). (1998). Effective teaching strategies that accommodate diverse learners. Columbus, OH: Merrill, Prentice Hall.

Kame'enui, E. J., & Simmons, D. C. (1990). Designing instructional strategies: The prevention of academic learning problems. Columbus, OH: Merrill.

Kaminski, R. A., & Good, R. H., III. (1996). Toward a technology for assessing basic early literacy skills. School Psychology Review, 25, 215-227.

Kaminski, R. A., & Good, R. H., III. (1998). Assessing early literacy skills in a problem-solving model: Dynamic indicators of basic early literacy skills. In M. R. Shinn (Ed.), Advanced applications of curriculum-based measurement. New York: Guilford.

Laimon, D. E. (1994). The effects of a home-based and center-based intervention on at-risk preschool children's early literacy skills. Unpublished Doctoral Dissertation, University of Oregon, Eugene.

LaBerge, D., & Samuels, S. (1974). Toward a theory of automatic information processing in reading. Cognitive Psychology, 6, 293-323.

Levy, B. A., Abello, B., & Lysynchuk, L. (1997). Transfer from word training to reading in context: Gains in reading fluency and comprehension. Language Disability Quarterly, 20, 173-188.

Linn, R. L. (2000). Assessments and accountability. Educational Researcher, 29(2), 4-16.

Logan, G. D. (1997a). Toward an instance theory of automatization. Psychological Review, 95, 492-527.

Logan, G. D. (1997b). Automaticity and reading: Perspectives from the instance theory of automation. Reading & Writing Quarterly, 13, 123-146.

Logan, G. (1988). Toward and instance theory of automatization. Psychological Review, 95, 492-527.

Lyon, R. (1997, July). Report on learning disabilities research at NIH. http://www.ldonline.org/ld_indepth/reading/nih_report.html, 1-11.

Markell, M. A., & Deno, S. L. (1997). Effects of increasing oral reading: Generalization across reading tasks. The Journal of Special Education, 31(2), 233-250.

Murnane, R. J., & Levy, F. (1996). Teaching the new basic skills: Principles for educating children to thrive in a changing economy. New York: The Free Press.

National Center for Education Statistics. (1999). NAEP 1998 reading: Report card for the nation and the states. Washington, D. C.: U. S. Department of Education, Office of Educational Research and Improvement.

National Reading Panel. (2000). Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction: Reports of the subgroups. Bethesda, MD: National Institute of Child Health and Human Development.

National Research Council. (1998). Preventing reading difficulties in young children. Washington, DC: National Academy Press.

Nunnally, J. C. (1978). Psychometric theory. (2nd ed.). New York: McGraw-Hill.

O'Connor, R. (2000). Increasing the intensity of intervention in kindergarten and first grade. Learning Disabilities Research and Practice, 15, 43-54.

Oregon Department of Education. (2000). Statewide assessment results 2000 [on-line]. Available: <http://www.ode.state.or.us/asmt/results/>.

Posner, M. I., & Snyder, C. R. R. (1975). Attention and cognitive control. In R. Solso (Ed.), Information processing and cognition: The Loyola Symposium (pp. 55-85). Hillsdale, NJ: Erlbaum.

Salvia, J., & Ysseldyke, J. E. (2001). Assessment. (8th ed.). Boston: Houghton Mifflin.

Shaywitz, S. E., Escobar, M. D., Shaywitz, B. A., Fletcher, J. M., & Makuch, R. W. (1992). Evidence that dyslexia may represent the lower tail of a normal distribution of reading ability. New England Journal of Medicine, 326, 145-150.

Shephard, L. A. (2000). The role of assessment in a learning culture. Educational Researcher, 29(7), 4-14.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic information processing: Perceptual learning, automatic attending, and general theory. Psychological Review, 96(84), 127-190.

Simmons, D. C., & Kame'enui, E. J. (Eds.). (1998). What reading research tells us about children with diverse learning needs: Bases and basics. Mahwah, NJ: Lawrence Erlbaum Associates.

Simmons, D. C., Kame'enui, E. J., & Good III, R. H. (1998). Accelerating children's competence in early reading and literacy-schoolwide: Project ACCEL-S (Federal OSEP grant #H324M980127): University of Oregon, Eugene, OR.

Simmons, D. C., Kame'enui, E. J., Good III, R. H., Harn, B. A., Cole, C., & Braun, D. (2000). Building, implementing, and sustaining a beginning reading model: School by school and lessons learned. Oregon School Study Council (OSSC) Bulletin, 43(3), 3-30.

Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. Reading Research Quarterly, 21, 360-406.

Stanovich, K. E. (2000). Progress in understanding reading: Scientific foundations and new frontiers. New York: The Guilford Press.

Tindal, G., Marston, D., & Deno, S. L. (1983). The reliability of direct and repeated measurement (Research Report No. 109). Minneapolis, MN: University of Minnesota Institute for Research on Learning Disabilities.

Thurlow, M. L., & Thompson, S. J. (1999). District and state standards and assessments: Building an inclusive accountability system. Journal of Special Education Leadership, 12(2), 3-10.

Torgesen, J. K. (1998). Catch them before they fall: Identification and assessment to prevent reading failure in young children. American Educator, 22(1), 32-39.

Wolf, M., Bowers, P. G., & Biddle, K. (2000). Naming-speed processes, timing, and reading: A conceptual review. Journal of Learning Disabilities, 33, 387-407.

Woodcock, R. W., & Johnson, M. B. (1989). Woodcock-Johnson Psychoeducational Battery-Revised. Allen: TX: DLM.

Yopp, H. K. (1995). Yopp-Singer test of phoneme segmentation. Newark, DE: International Reading Association.

ACADEMY GUIDEBOOK

This publication was developed by RMC Research Corporation and was funded by The Partnership for Reading, a project administered by the National Institute for Literacy. The Guidebook is based on products and materials created by consultants, researchers, and technical assistance providers for the Secretary's Reading Leadership Academies, sponsored by the U. S. Department of Education and The Partnership for Reading, in January and February, 2002.

The Partnership for Reading is a collaborative effort of the National Institute for Literacy, the National Institute of Child Health and Human Development, the U.S. Department of Education, and the U.S. Department of Health and Human Services to make evidence-based reading research available to educators, parents, policy makers, and others with an interest in helping all people learn to read well. The findings and conclusions in this publication are based on the 2000 report of the National Reading Panel, *Teaching Children to Read: An Evidence-Based Assessment of the Scientific Research Literature on Reading and Its Implications for Reading Instruction—Report of the Subgroups*.

The National Institute for Literacy, an independent federal organization, supports the development of high-quality state, regional, and national literacy services so that all Americans can develop the literacy skills they need to succeed at work, at home, and in the community.

The findings and conclusions in this publication do not necessarily represent the positions or policies of the U.S. Department of Education or The Partnership for Reading.

Editor

C. Ralph Adler

Designer

Diane Draper

Additional contributors at RMC Research Corporation

Jan Merkley, Shawna Godbout, Richard O'Rourke, Bob Kozman, Susan Klaiber

To order additional copies, contact the National Institute for Literacy at EdPubs, PO Box 1398, Jessup, MD 20794-1398. 800-228-8813.

The Reading Leadership Academy Guidebook can also be downloaded at The Partnership for Reading web site, www.nifl.gov/partnershipforreading/.

August 2002



National Institute for Literacy

National Institute of Child Health
and Human Development

U.S. Department of Education

U.S. Department of Health and Human Services



National Institute for Literacy

National Institute of Child Health
and Human Development

U.S. Department of Education

U.S. Department of Health and Human Services

September 2002